SYSTEM IDENTIFICATION FROM NONUNIFORMLY SAMPLED DATA

A thesis submitted
in partial fulfilment of the requirements
for the degree of

MASTER OF TECHNOLOGY

By

ABHAY C. RANADE

to the

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

FEBRUARY, 1990.

CERTIFICATE

It is certified that this work entitled ***SYSTEM IDENTIFICATION FROM NONUNIFORMLY SAMPLED DATA*** *by* ***ABHAY C RANADE*** *has been completed under my supervision and that this work has not been submitted elsewhere for the award of a degree*

Pradip Sircar
5/2/1990

Dr P Sircar
Assistant Professor
Dept of Electrical Engineering
Indian Institute of Technology,
KANPUR - 208016, INDIA

EE- 1990-M-RAN-SYS

ACKNOWLEDGEMENT

It is my pleasant duty to express my sincere gratitude to Dr P Sircar his encouragement, valuable comments and constructive suggestions I would lik thank him for his many ideas which helped me to broaden my perspective on si processing I express my appreciation to him for guiding my efforts and shar my enthusiasm in this challenging project

No words are adequate to express by indebtedness to my mother for all pains and sufferings she has undergone to bring me up to this stage

DEDICATED TO

AIEE AND ATUL

ABSTRACT

In this thesis, a technique has been presented for system identification from a nonuniformly sampled data In this regard, the idea of Orthogonal Polynomial Approximation is used for signal reconstruction Accurate parameter estimation is achieved by using the statistical properties of the colored noise in the reconstructed signal The estimation procedure presented here makes use of the subspace separation approach Singular Value Decomposition (SVD) has been used, in this regard Furthermore, the ever present problem of model order selection has been addressed adequately

The technique presented here, was also applied to the special case of uniform sampling Performance comparison, with a conventional technique, for the case of uniformly sampled data, proved the superiority of the technique given here

CONTENTS

PAGE No

CHAPTER 1
INTRODUCTION

Accurate estimation of parameters of a signal embedded in noise is a problem faced in many signal processing applications Particularly important is the case when the parameters to be estimated are time-invariant These parameters are estimated from a given set of measurements of a signal These measurements can be performed in two ways, viz , by uniform sampling or by nonuniform sampling In this thesis we consider the case of nonuniform sampling

Nonuniform sampling can be chosen over uniform sampling for various reasons Efficiency of sampling and noise reduction are two such considerations favouring nonuniform sampling

An efficient sampling strategy would be one which has a capability of adjustable sampling rate The sampling rate is adjusted, according to the rate of change of signal Thus, more number of samples are taken in the region where the signal changes rapidly, and reduced number of samples in the region where the signal varies slowly

Yet another sampling strategy will be to use nonuniform sampling to increase the SNR of the sampled points This can be achieved for a signal embedded in zero mean noise, by sampling more number of points when the signal is away from zero crossing, and by sampling less number of points when the signal is near the zero crossing The reasoning behind the strategy should be apperent

In both of the sampling schemes, it is evident that the sampling points will be distributed arbitrarily, although it will be necessary that the *sampling theorem* be satisfied *locally* [1] in order to achieve complete reconstruction of the signal

Modern signal processing methods use parametric modeling to achieve high resolution and accuracy The idea is to choose a particular model depending on a priori information about the signal Use of incorrect model, however, deteriorates the overall performance

Modeling of an ARMA process by an AR process is a typical example, where use of incorrect model reflects in certain fundamental limitations like high model order requirement and increased computational burden Another example is of

parameter estimation of a signal consisting of damped sinusoids Use of adaptive spectral estimation techniques in this case has been suggested In such an attempt, however, the model itself becomes *time-varying* Thus, it becomes necessary to adapt the model parameters to the *nonstationarity* in the given signal Needless to say, such methods are restricted to only those applications where a large record of data is available, and also the damping factors are not too large

In this thesis, we present a *time-invariant* model to characterise the signal consisting of *complex exponentials* embedded in noise Thus, our requirement is to estimate a constant set of parameters accurately, instead of fast and reliable tracking of time-varying parameters Note that the use of small data records for parameter estimation is our greatest constraint, and under this constraint we will fully utilize the nonuniform sampling scheme, and achieve efficient sampling together with high accuracy in estimation However, throughout this thesis we will maintain arbitrariness of the distribution of sampled points, except that the condition for complete reconstruction of signal is satisfied

Sircar and Sarkar [1] have worked on system identification from a nonuniformly sampled signal It has been shown that the estimated poles are accurate only for high SNR, but they deteriorate substantially as the SNR is reduced, a factor which is attributed to the numerical ill-conditioning of the problem itself Neverthless, the fact that their model dose'nt take into account the embedded noise should not be overlooked Thus, by utilizing information about the noise while estimating the signal poles, it can be expected that substantial improvement in accuracy can be obtained

In this thesis, we make use of the so called *Orthogonal Polynomial Approximation*, as used by Sircar and Sarkar [1], for reconstruction of nonuniformly sampled signal In the present work, we will also characterise the noise present in this approximating (reconstructed) signal and use that information to get accurate estimates of the signal parameters

It was observed in the process of simulation, that the noise present in the approximating signal was colored, even when the noise embedded in the signal was white

A colored noise sequence is characterised by the fact that it is correlated, unlike white noise sequence, which is uncorrelated Thus, colored noise is a *nondeterministic, bandlimited* signal

A major contribution of this thesis is that a technique has been suggested which can estimate the signal parameters even in the presence of colored noise at the first place In this regard, we assume that the so called *Normalised Covariance Function* of the noise, embedded in the nonuniformly sampled signal, is known a priori For the purpose of simulation we assumed the noise to be *white, Gaussian* and *Wide Sense Stationary* (WSS)

It is interesting to note that the colored noise in the approximating signal originates because of the polynomial approximation procedure Note that the white noise assumption is not stringent, since, once the Normalised Covariance Function is known, both the cases, viz , white or colored, become exactly same

A variety of applications which can use the above mentioned methodology, are given below

In Electro Magnetic Pulse (EMP) applications [2], one of the primary requirements is to obtain the transient response of a test object In this scheme, an Electro Magnetic Pulse is impinged on the test object The output of the object is sensed by a probe, having a known transfer function, and is digitised These samples are then used to identify the transfer function of the test object A typical EMP has a short rise time and rather slow decay rate Thus, nonuniform sampling will certainly be benificial in this case

Another application could be in high resolution *Direction Of Arrival* (DOA) estimation [3] DOA estimation is important in many sensor systems such as radar, sonar, electronic surveillance and seismic exploration The transmission medium is assumed to be *isotropic* and *nondispersive* so that the radiation propogates in straight lines, and the sources are assumed to be in the far-field of the array We know that if and only if the transmission medium is vacoum, the signal propogates undamped But in applioations like radar, sonar and seismio exploration, the signal could be sufficiently damped, thereby making it nonstationary [4] Algorithms like *MUSIC* (MUltiple SIgnal Classifioation) [5] and *ESPRIT* (Estimation of Signal Parameters via Rotational Invariance Teohnique) [6]

have been suggested for solving the above mentioned problem MUSIC assumes that the signal is stationary, which may not be the case in true sense Also, both MUSIC and ESPRIT use uniform samples and assume that the embedded noise is white The later assumption is not true in most of the practical cases

A very important application is in *Linear Predictive Coding* (LPC) [7] for bandwidth reduction of speech signals The idea is to use an AR model to represent the speech spectrum Assuming that speech signals can be accurately modeled as the output of an all pole filter driven by white noise for unvoiced speech, or driven by train of impulses for voiced speech, the speech waveform may be reduced to a small set of parameters Thus, only the model parameters and the period of the impulse train need be transmitted or stored Speech synthesis is achieved using appropriate model for each speech sound

Note that we have considered the case where the sampling grid is not fixed but is arbitrary In a fixed sampling grid strategy, a globally optimum sampling grid is searched for and is used to sample the given signal In contrast to this, we have considered the case where the sampling grid structure changes with the signal, depending upon the efficiency and/or noise reduction considerations

Distributed Radar Detection Systems (DRDS) [7] make use of a fixed nonuniform sampling grid strategy The idea is to find globally optimum struotures and strategies for various elements and then logically combine all the respeotive signals in the so called central processor It is observed that the optimum peripheral detector thresholds depend upon the central processor strategies, (AND, OR, etc) and vice versa

CHAPTER 2
FORMULATION OF THE PROBLEM AND MEASUREMENT INTERPOLATION

2.1 *INTRODUCTION* -

In this chapter we discuss the problem of transforming a nonuniformly sampled data to uniformly sampled one, using *orthogonal polynomial approximation* This step will be necessary in order to be able to apply the parameter estimation techniques given in the subsequent chapters We also discuss, the effect of noise embedded in the nonuniformly sampled data, on the approximating signal

We have considered the more general case of *complex exponentials* It may, however, be noted that the measurement interpolation procedure itself is independent of the actual signal used Lastly we will discuss the relationship between the noise in the approximating signal to that of the nonuniformly sampled signal

2.2 *FORMULATION* -

A complex signal r(t) which consists of M complex exponentials, can be expressed as,

$$r(t) = \sum_{i=1}^{M} A_i e^{s_i t} \quad , \; t \geq 0 \qquad (2.2.1)$$

For a real signal, the expression becomes a sum of damped sinusoids and we will have one complex conjugate pair of roots in the s domain for each of the damped sinusoids

Let y(t) be the signal which is equal to r(t) added with corrupting noise w(t) Thus , y(t) can be written as,

$$y(t) = r(t) + w(t)$$

$$= \sum_{i=1}^{M} A_i e^{s_i t} + w(t) \quad , \; t \geq 0 \qquad (2.2.2)$$

Let (t_k k = 1, 2, K), be the time instants where the signal y(t) is sampled Note that these time instants need not be at uniform spacing Uniform sampling becomes a special case of the given problem Thus, the corrupted signal y(t), at the instants t_k, will be given by,

$$y(t_k) = \sum_{i=1}^{M} A_i e^{s_i t_k} + w(t_k) \qquad (2.2.3)$$

We have assumed $w(t_k)$ to be a complex, zero mean, Gaussian stationary sequence with known *normalised autocovariance function* White noise assumption for $w(t_k)$ will be a special case where the normalised covariance function will be unity for lag zero, but will be zero elsewhere Also, as $w(t_k)$ is assumed to be a *stationary process*, its normalised autocorrelation matrix will be *symmetric* and have a *Toeplitz structure*

Thus, our problem is to estimate the parameters of r(t) using the corrupted samples $y(t_k)$

2.3 *ORTHOGONAL POLYNOMIAL APPROXIMATION* -

Our first job will be to transform the nonuniform samples $y(t_k)$, to a signal x(t), which will approximate y(t) over the sampling span If the sampling theorem is satisfied locally [1] then complete reconstruction of the original signal is possible We will be using orthogonal polynomials as given in [1] to generate the approximating signal x(t) Thus, x(t) will be given by,

$$x(t) = \sum_{j=1}^{N} d_j p_{j-1}(t), \qquad (2.3.1)$$

where N is the order of the approximating polynomial

For any arbitrary set of sampling grid (t_k k = 1, 2, K), we can evaluate the orthogonal polynomials p_j from the recurrence relations given by,

$$p_j = (t - a_j) p_{j-1} - b_{j-1} p_{j-2}, \qquad (2.3.2)$$
$$\text{with } p_0 = 1, p_{-1} = 0,$$

&

$$a_j = \frac{1}{\phi_{j-1}} \sum_{k=1}^{K} t_k \, [\, p_{j-1}(t_k)]^2,$$

$$b_j = \frac{\phi_j}{\phi_{j-1}} ,$$

$$\phi_j = \sum_{k=1}^{K} [\, p_j(t_k) \,]^2$$

It should be noted that the this orthogonal set of polynomials will be independent of the actual value of $y(t_k)$ but will be dependent only on the nonuniform sampling grid Once the orthogonal polynomials p_j are calculated, we can also find d_j To calculate d_j we make use of equation (2 3 1) by replacing x(t) in it, by y(t), at t_k

Therefore,

$$y(t_k) = \sum_{j=1}^{N} d_j \, p_{j-1}(t_k) \qquad (2\ 3\ 3)$$

Rewriting equation (2 3 3) in vector form, we have,

$$Y = PD \qquad (2\ 3\ 4)$$

where, Y is the input vector given by,

$$Y = [\, y(t_1) \;\; y(t_2) \;\; \ldots \;\; y(t_K) \,]^T \qquad (2\ 3\ 5)$$

P is given by,

$$P = \begin{bmatrix} p_0(t_1) & p_1(t_1) & \ldots & p_{N-1}(t_1) \\ p_0(t_2) & p_1(t_2) & \ldots & p_{N-1}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ p_0(t_K) & p_1(t_K) & \ldots & p_{N-1}(t_K) \end{bmatrix} \qquad (2\ 3\ 6)$$

and D is the coefficient vector, given by,

$$D = [\, d_1 \quad d_2 \quad \cdots \quad d_N \,]^T \qquad (2.3.7)$$

Equation (2.3.4) can be solved for D, as both Y and P are known

We now elucidate how the orthogonality property of the approximating polynomials can be used to find D Multiplying both sides of equation (2.3.4) by P^T, we get,

$$P^T Y = P^T P D \qquad (2.3.8)$$

Equation (2.3.8) can be expressed as,

$$P^T Y = \begin{bmatrix} \sum_{k=1}^{K} p_0^2(t_k) & 0 & \cdots & 0 \\ 0 & \sum_{k=1}^{K} p_1^2(t_k) & & \\ \vdots & & \ddots & 0 \\ 0 & & 0 & \sum_{k=1}^{K} p_{N-1}^2(t_k) \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix}$$

(2.3.9)

From equation (2.3.9), we observe that $P^T P$ is a diagonal matrix This is because of the orthogonality property of the approximating polynomials The orthogonality property states that,

$$\sum_{k=1}^{K} p_i(t_k)\, p_j(t_k) = 0, \qquad \text{for } i \neq j$$

$$= \text{nonzero}, \qquad \text{for } i = j \qquad (2.3.10)$$

We now multiply both sides of equation (2.3.8) by $(P^T P)^{-1}$ Thus, we get,

$$D = (P^T P)^{-1} P^T Y \qquad (2\ 3\ 11)$$

As $P^T P$ is a diagonal matrix, it is evident that the inverse of $P^T P$ is also a diagonal matrix, where each of the diagonal element is reciprocal of the corresponding element of $P^T P$ Thus, the choice of orthogonal polynomials effects in reducing the computational burden required to solve equation (2 3 11)

Once d_j are found, these can be substituted in the equation (2 3 1) to find x(t)

Let $e(t_k)$ be the difference between $y(t_k)$ and $x(t_k)$
Therefore,

$$e(t_k) = y(t_k) - x(t_k) \qquad (2\ 3\ 12)$$

Let $e(t_k)$ be the approximating error Thus, $e(t_k)$ will be the difference between $x(t_k)$ and $r(t_k)$
Thus, we get,

$$e(t_k) = x(t_k) - r(t_k) \qquad (2\ 3\ 13)$$

Therefore,

$$x(t_k) = r(t_k) + e(t_k) \qquad (2\ 3\ 14)$$

We observe, from equation (2 3 14), that the approximating signal $x(t_k)$ consists of the original signal $r(t_k)$ and the approximating error $e(t_k)$ Also, from equation (2 3 12), it follows that,

$$y(t_k) = x(t_k) + e(t_k) \qquad (2\ 3\ 15)$$

Substituting $x(t_k)$ from equation (2 3 14) in equation (2 3 15), we get,

$$y(t_k) = r(t_k) + [\ e(t_k) + e(t_k)\ ] \qquad (2\ 3\ 16)$$

Comparing equations (2 2 3) and (2 3 16), we get,

$$w(t_k) = [\ e(t_k)\ ] + [\ e(t_k)\ ] \qquad (2\ 3\ 17)$$

Note that the SNR, as mentioned here, is the peak SNR This means that the SNR is calculated by considering the peak of the signal r(t) It follows that the *local SNR* will go on decreasing as the signal damps out

EXAMPLE -

r(t) is given by,

$$r(t) = a_1 e^{s_1 t} + a_2 e^{s_2 t} + a_3 e^{s_3 t} + a_4 e^{s_4 t},$$

where, $a_1 = a_2 = a_3 = a_4 = 1.5$,

$s_1 = -0.01322 + j\,2\pi 0.04$, $s_2 = -0.01322 - j\,2\pi 0.04$,

$s_3 = -0.017664 + j\,2\pi 0.05456$, $s_4 = -0.017664 - j\,2\pi 0.05456$

SNR = 5 dB

Fig 2.1 shows the plot for $r(t_k)$ and $x(t_k)$. Note that the samples for $r(t_k)$ are nonuniform and arbitrary to the extent that the maximum inter sample distance does not fall below the minimum *local Nyquist rate* required.

Fig 2.2 shows the plot for $w(t_k)$ and $e(t_k)$. A very important conclusion can be drawn from this graph. Looking at the nature of $e(t_k)$ we can see that it is not white, but is bandlimited.

From Fig 2.3 we see that $c(t_k)$ also is not white, but is bandlimited. Thus, by orthogonal polynomial approximation, we have effectively split the spectrum of $w(t_k)$ in a upper frequency portion, which is $c(t_k)$, and a lower frequency portion, which is $e(t_k)$. The frequency at which this seperation takes place is dependent on the polynomal order selected. It is interesting to see from Fig 2.2 that $e(t_k)$ is approximately equal to the average of $w(t_k)$, where the averaging process takes into account only a limited number of samples around a particular point. Apparently, this number depends on the polynomial order selected. Fig 2.2 is for polynomial order (N) of 22.

Fig 2.4 shows the graph of $w(t_k)$ and $e(t_k)$, but with N = 19. It can be easily seen from Fig 2.4 that $e(t_k)$ is smooth as compared to $e(t_k)$ in Fig 2.2. Fig 2.3 and Fig 2.5 show $c(t_k)$ for N = 22 and N = 19 respectively.

From Fig 2.2 and Fig 2.4, we can conclude that $e(t_k)$ is present because of the low frequency components in $w(t_k)$, which in turn introduces error in $x(t_k)$. To support this statement we copy all the data points of $c(t_k)$ in $w(t_k)$ so that $w(t_k)$ becomes bandlimited. Now we perform the polynomial approximation procedure with

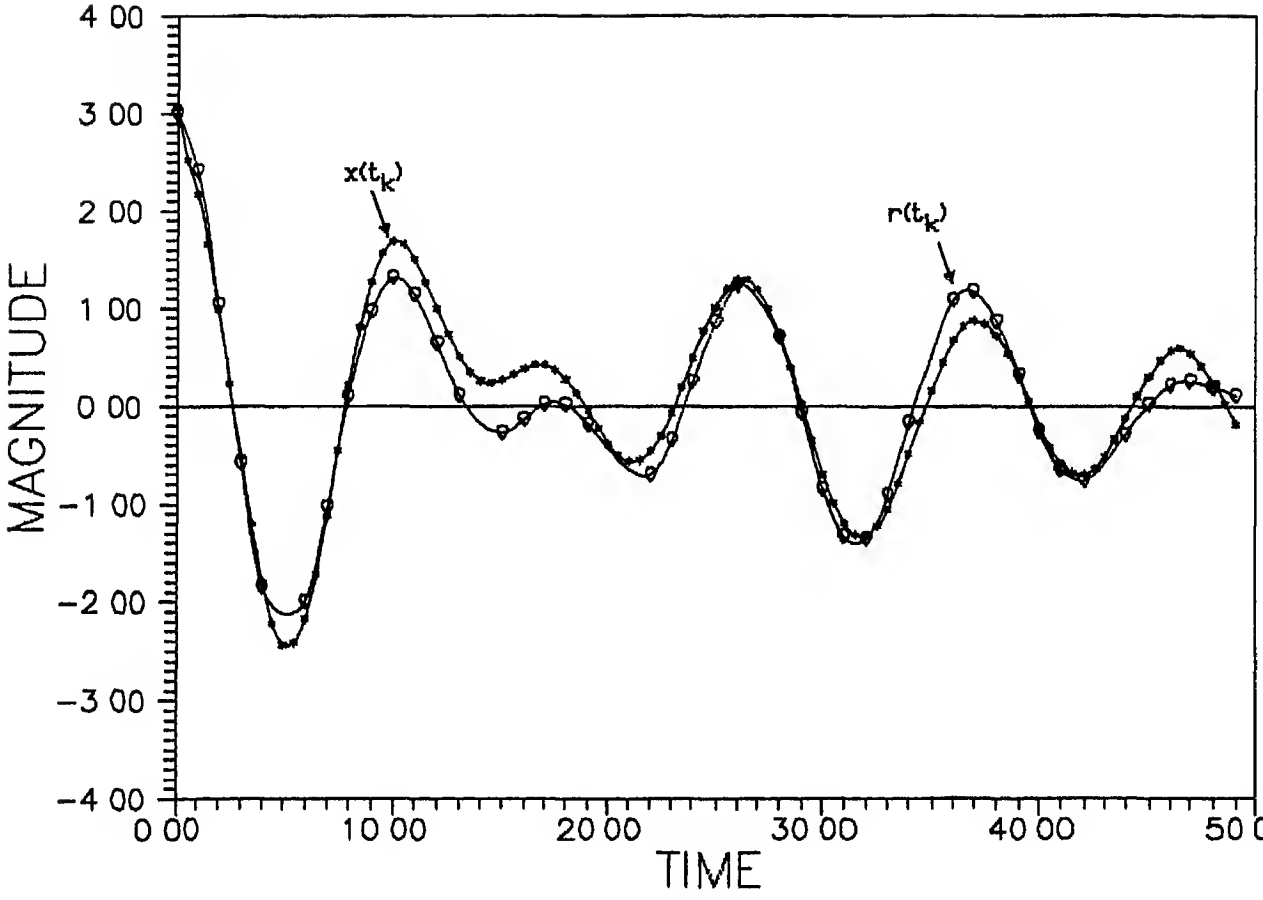

Fig 21

SNR = 5 dB. (White noise.) Polynomial order = 22

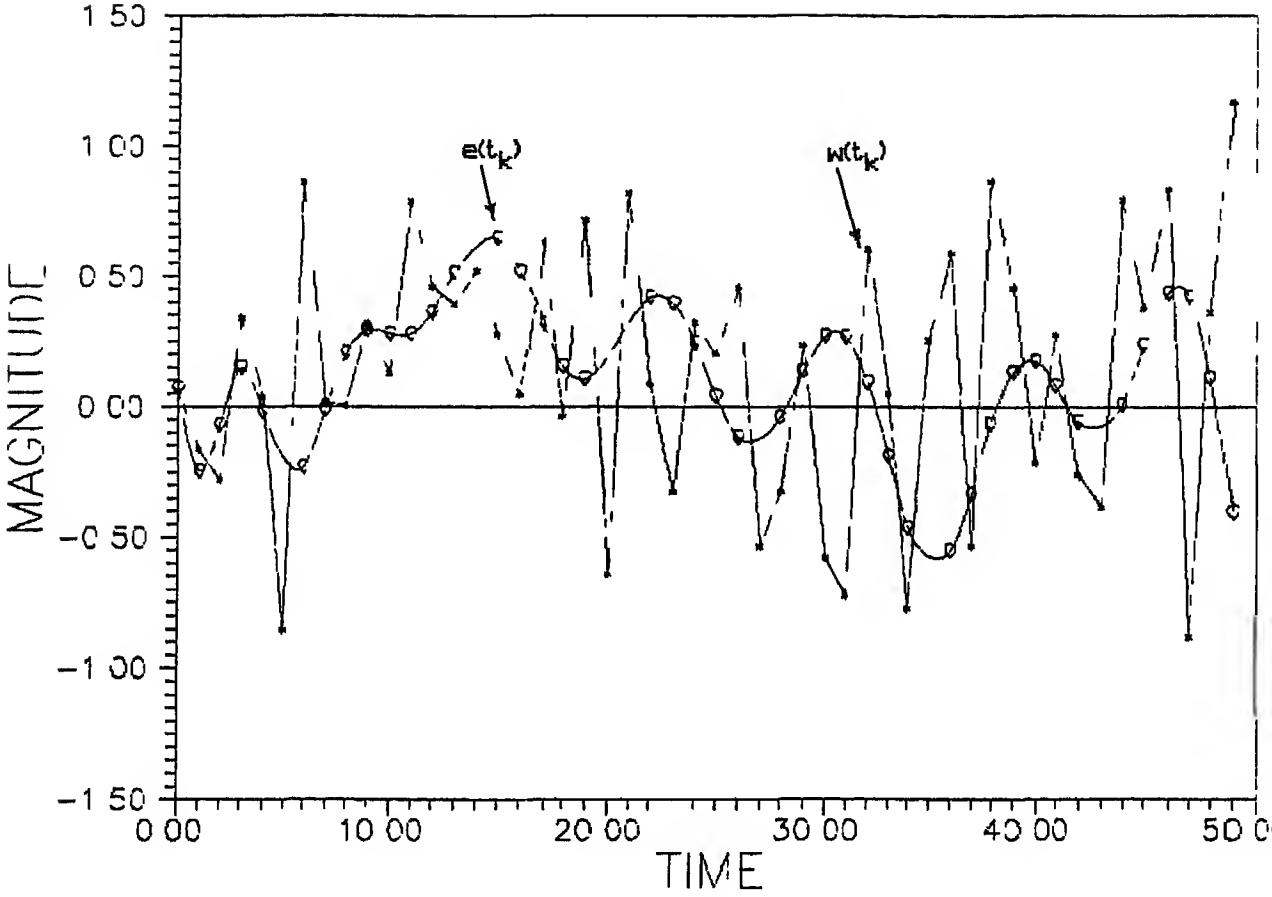

Fig 2.2

SNR = 5 dB. (White noise) Polynomial order = 22.

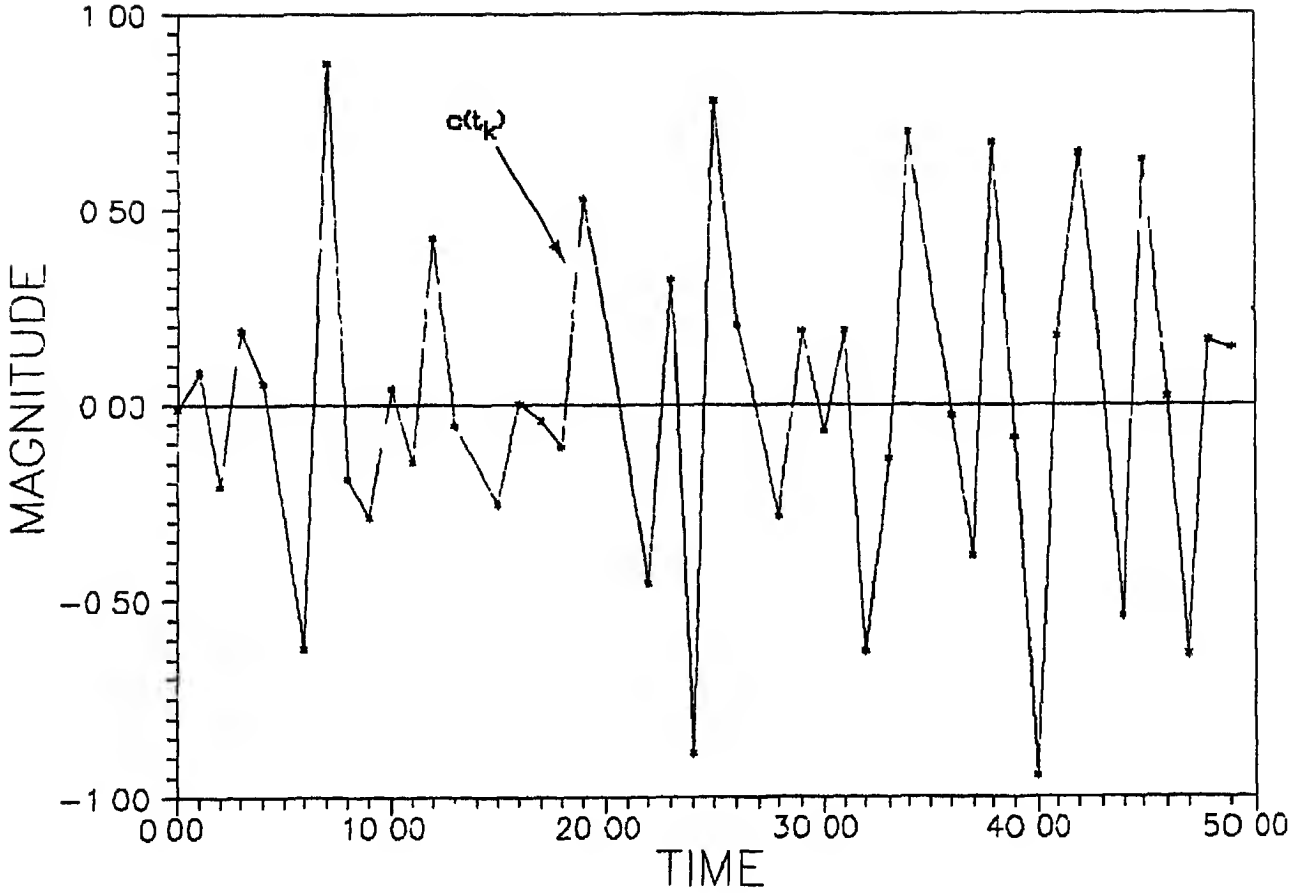

Fig 2.3.

SNR = 5 dB. (White noise.) Polynomial order = 22.

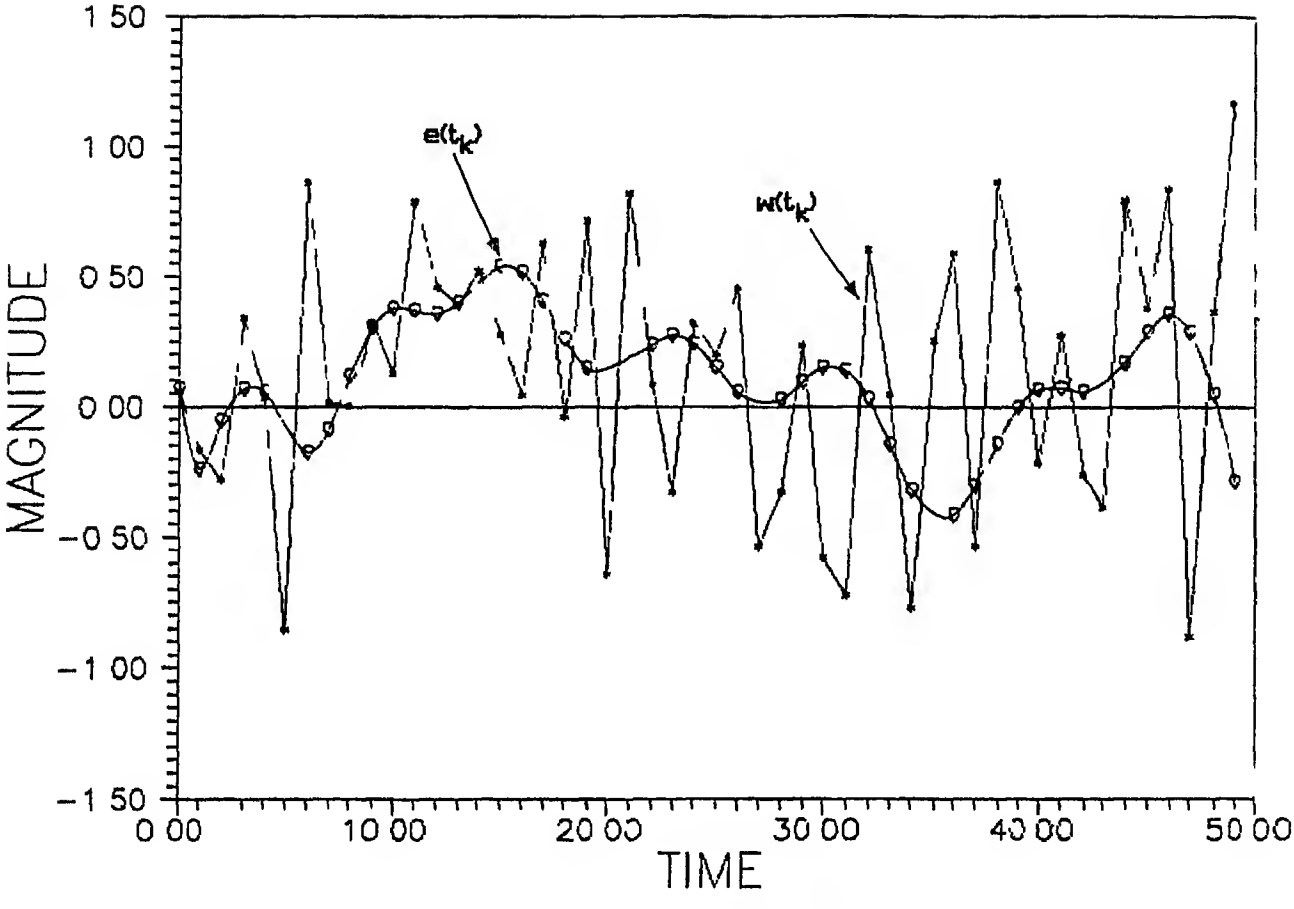

Fig. 2.4.

SNR = 5 dB. (White noise.) Polynomial order = 19.

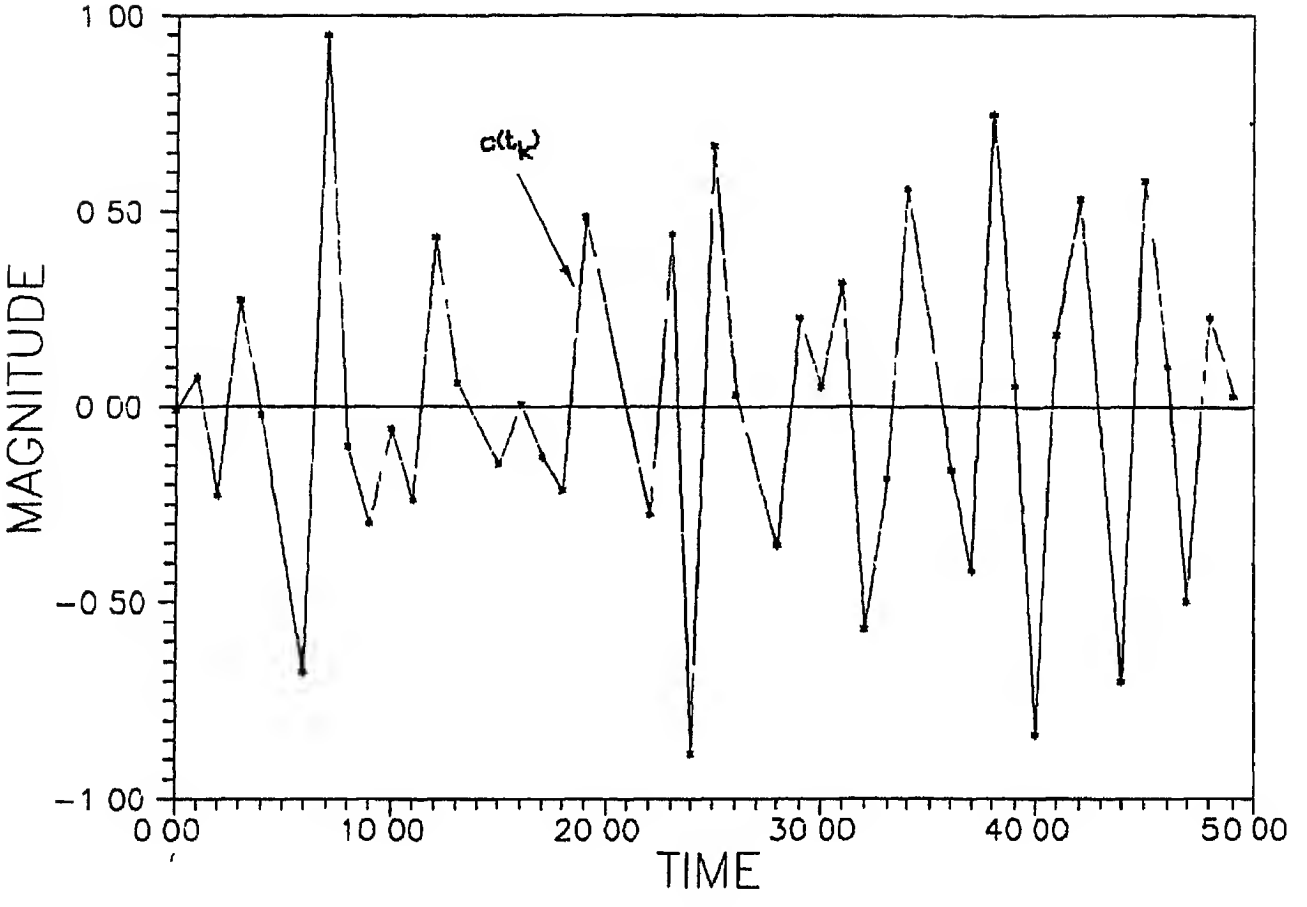

Fig 2.5

SNR = 5 dB. (White noise.) Polynomial order = 19.

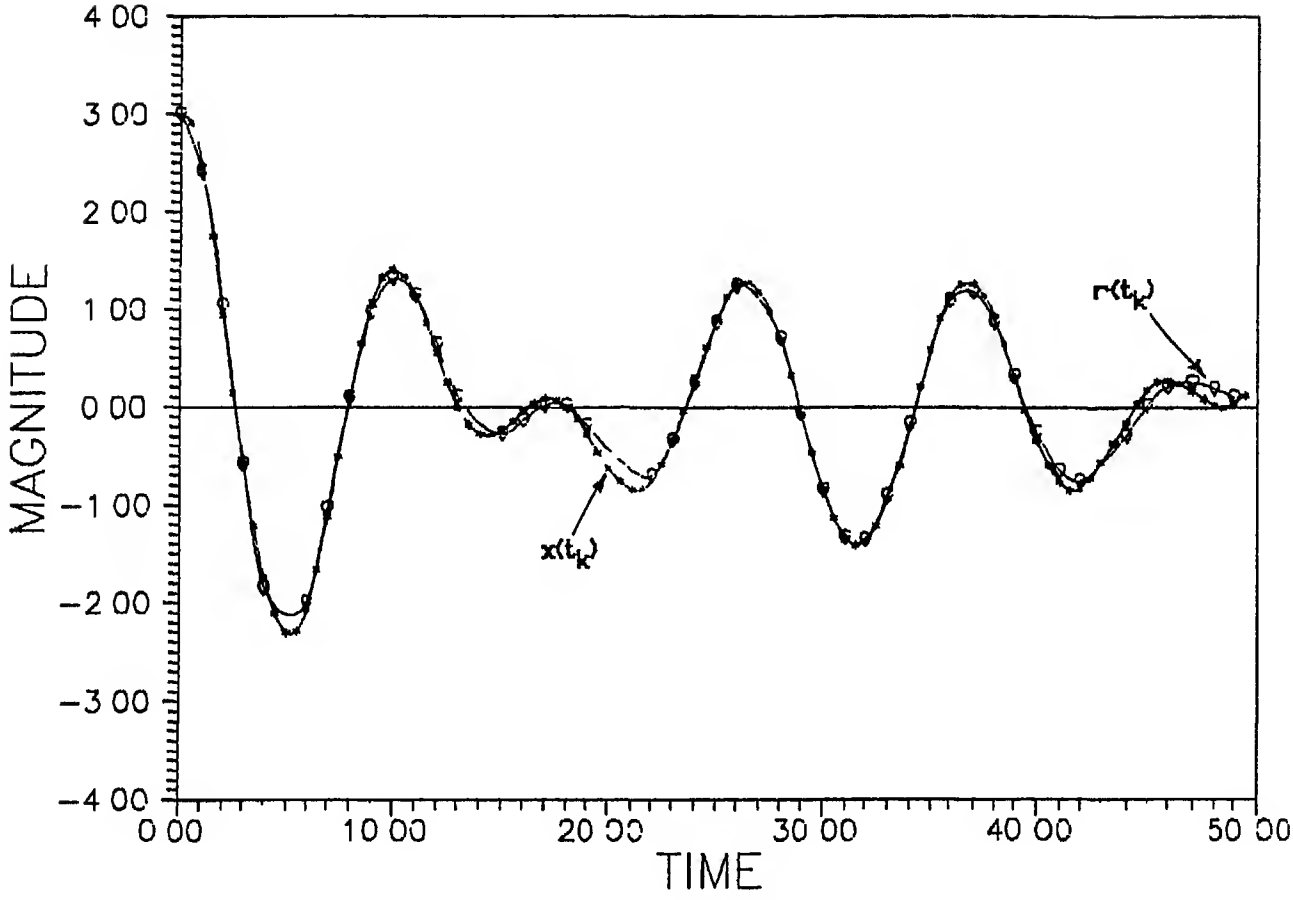

Fig 2.6

SNR = 5 dB. (Colored noise) Polynomial order = 21.

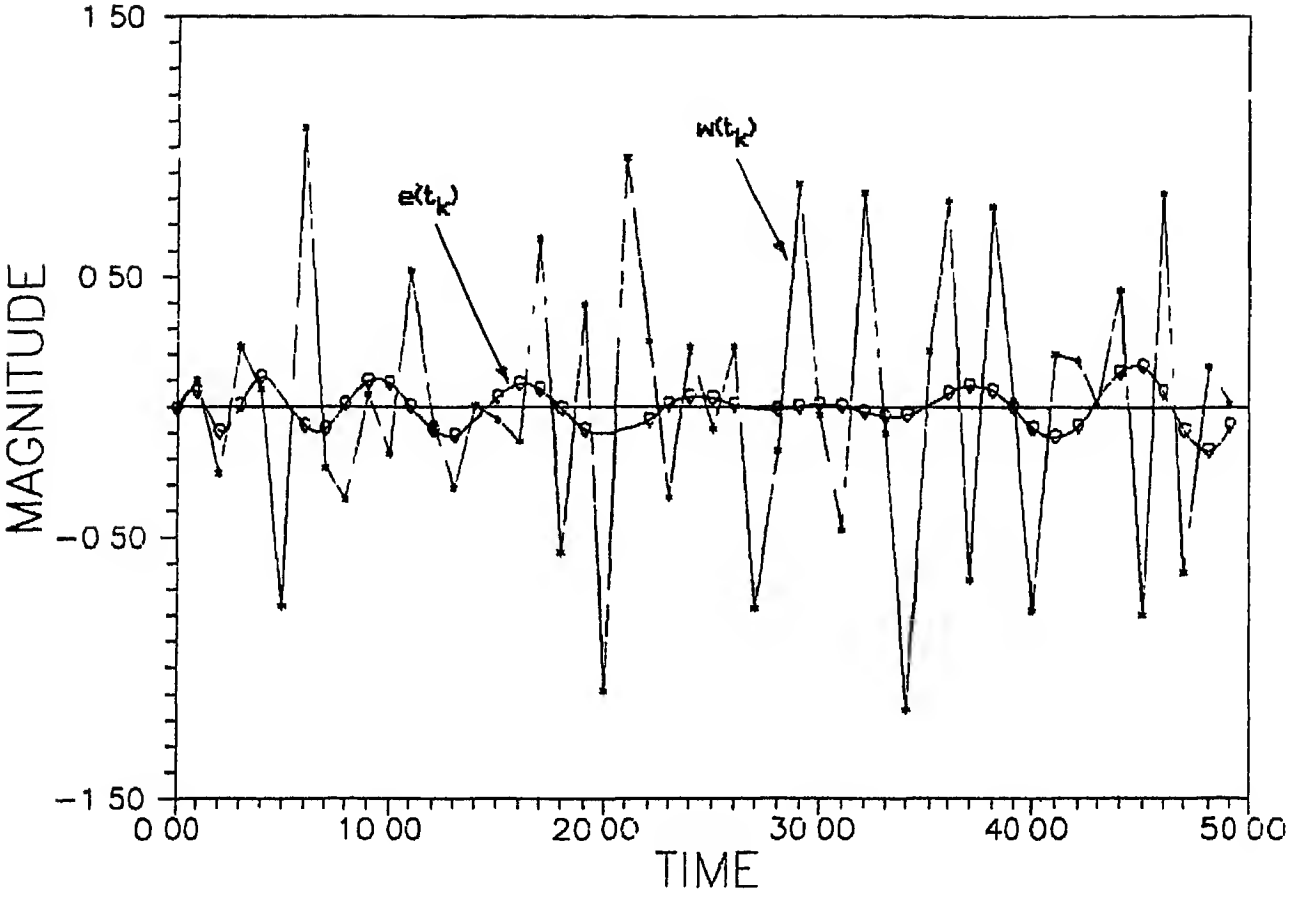

Fig 27

SNR = 5 dB. (Colored noise.) Polynomial order = 21.

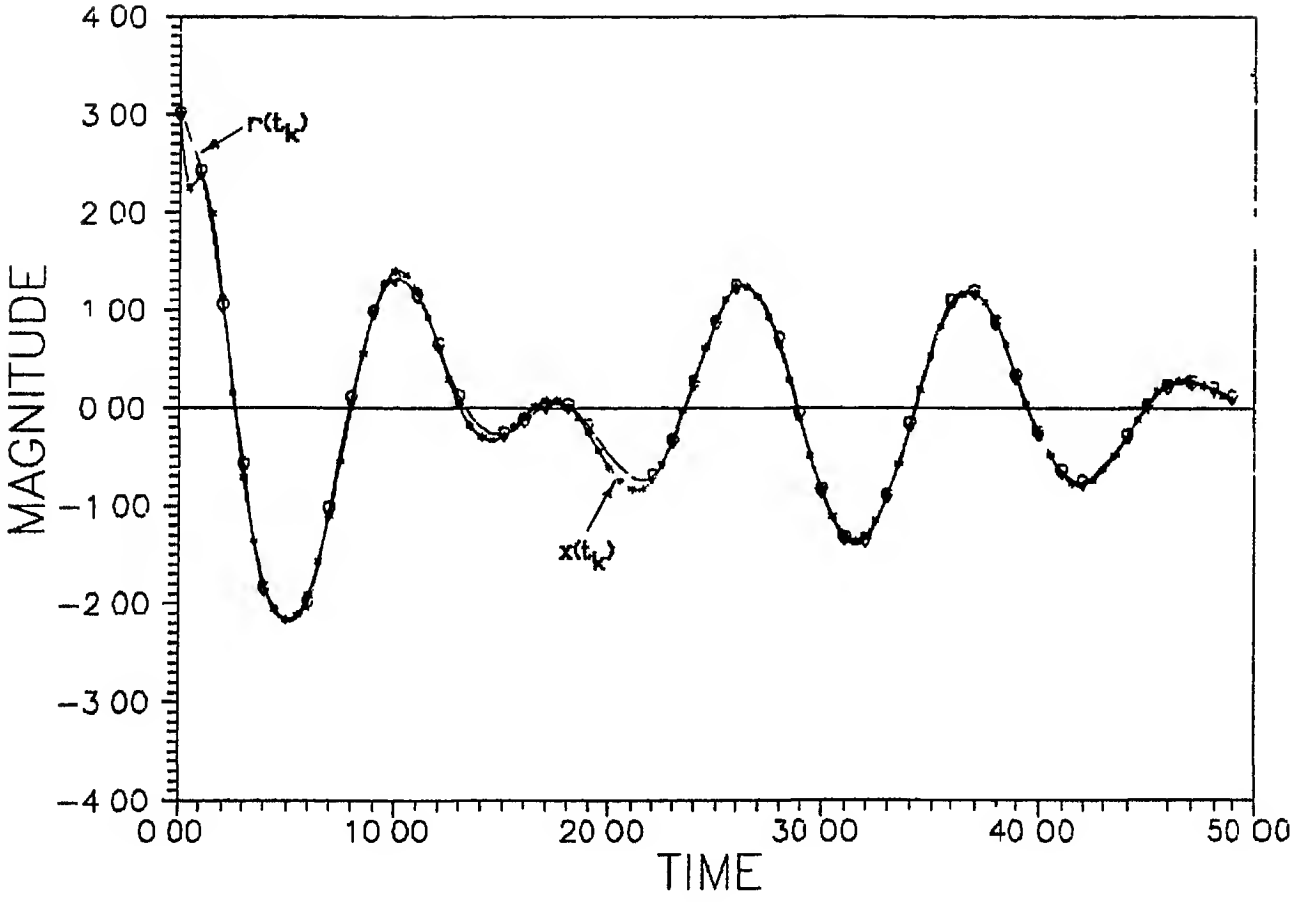

Fig 28

SNR = 40 dB. (White noise.) Polynomial order = 22.

SNR of 5 dB Fig 2 6 shows the plot of $r(t_k)$ and $x(t_k)$ in this case Note that $r(t_k)$ has remained same but $x(t_k)$ has changed drastically as compared to the initial case of white noise, and is almost equal to $r(t_k)$ This is evident in Fig 2 7 which shows the plot of $w(t_k)$ and $e(t_k)$ From Fig 2 7 we see that $e(t_k)$ has a small variation compared to the variation in Fig 2 4 This is because of the absence of the low frequency components in $w(t_k)$

Fig 2 8 shows the plot for $r(t_k)$ and $x(t_k)$ for the same case but with SNR of 40 dB From these plots, we can see that the approximating error is negligible

2 4 SELECTION CRITERIA FOR THE ORDER OF THE APPROXIMATING POLYNOMIAL -

We will use the error variance as a measure for selecting the model order The error varianoe is defined [1] as,

$$\sigma'^{2}_{w} = \frac{\sum_{k=1}^{K} [\, y(t_k) - x(t_k) \,]^2}{K - N} \qquad (2\ 4\ 1)$$

where K is the number of nonuniformly sampled points and N is the order of the approximating polynomial

There can be two criteria for selecting the order of the approximating polynomial

(i) Select that order which gives a minimum error variance
(ii) Select that order after which the error variance does not decrease appreciably

It was observed from simulation that if $r(t_k)$ itself has rapid variations in it, then both of the above mentioned criteria select the same order But if $r(t_k)$ is smooth enough then the order selected through the first criterion is generally large as compared to the order selected through the second criterion It is also observed, that the amount of increase in error variance, is not appreciable, although the model order is reduced, while using the second criterion Thus, to save further computational burden, one should opt for the second criterion

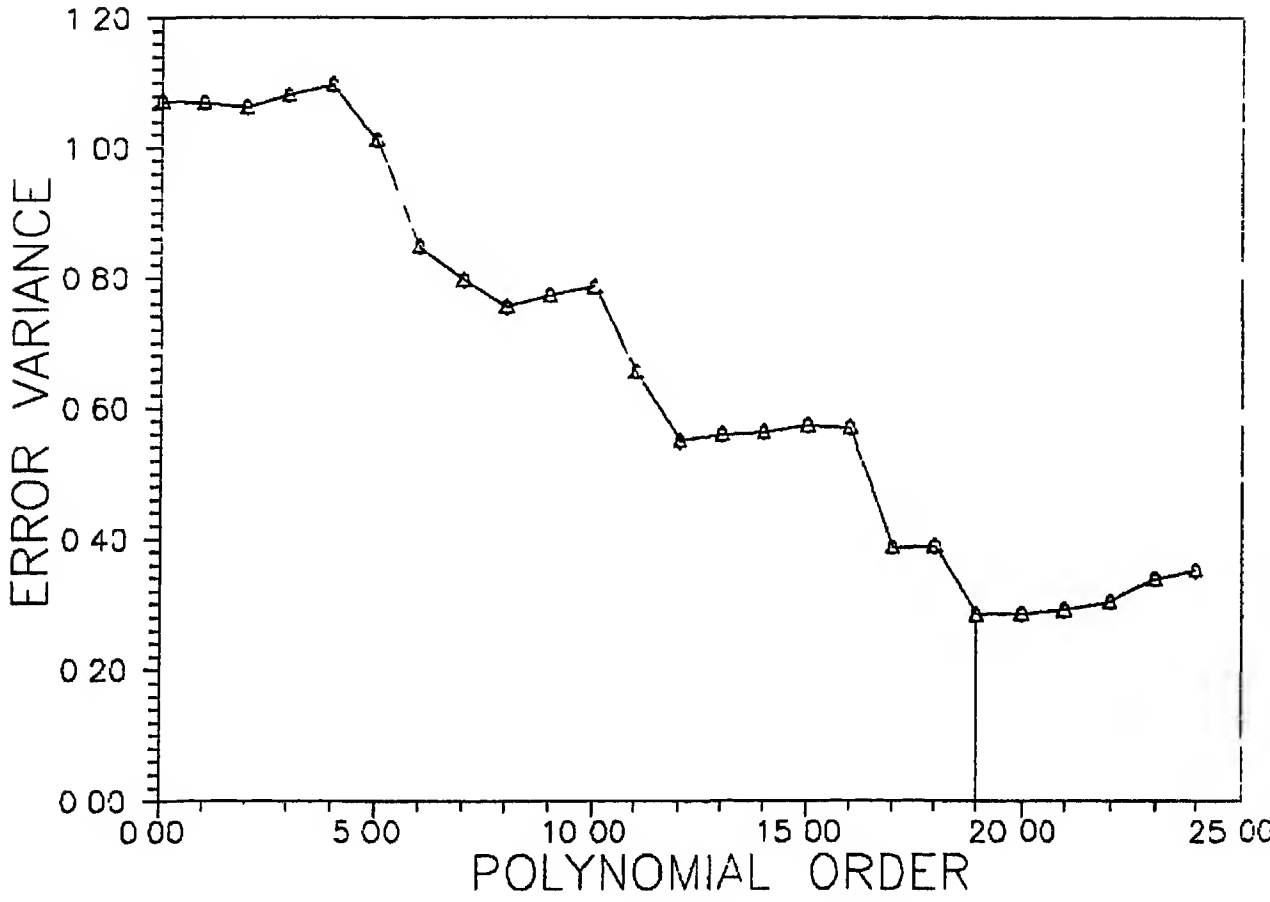

Fig 29

SNR = 5 dB. (White noise.)

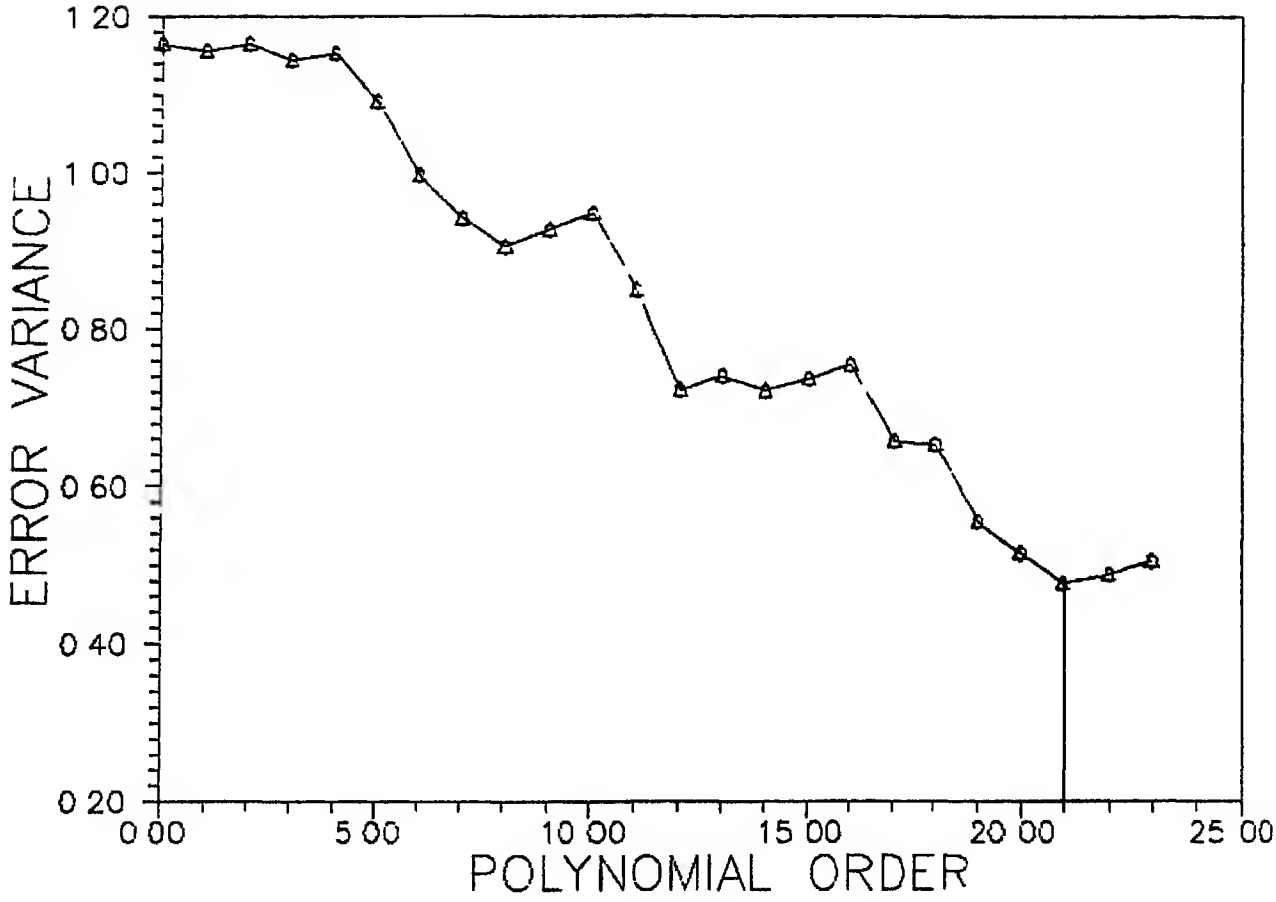

Fig 2.10

SNR = 5 dB. (Colored noise.)

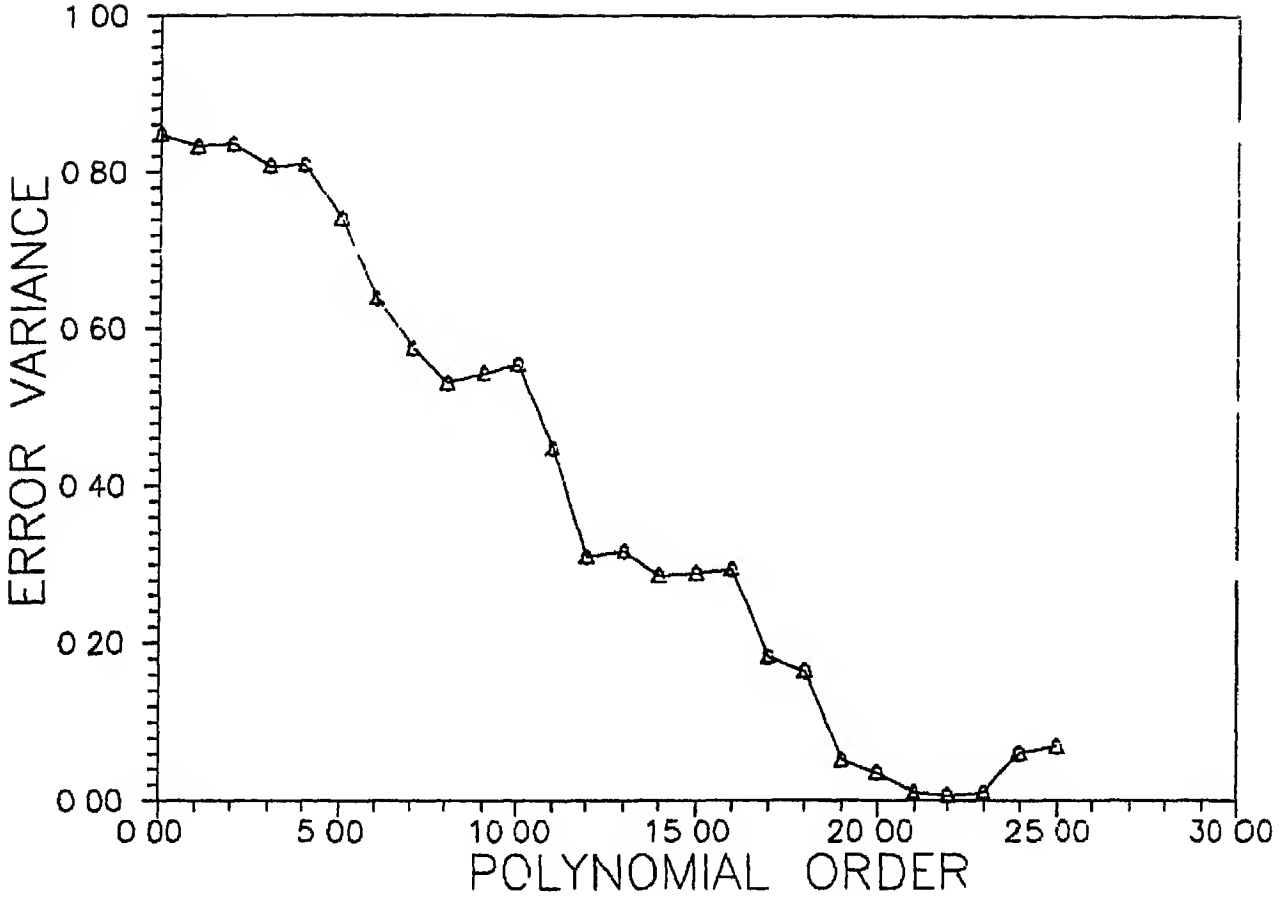

Fig. 2.11

SNR = 40 dB. (White noise.)

It should be noted that, to find the correct model order, we have to find the error variance for all the model orders, starting from 1, to the maximum value possible (which is decided by the machine capacity) After this, select the appropriate order, and repeat the approximation procedure for that order only

Fig 2 9 shows the plot of error variance vs polynomial order for a sequance $w(t_k)$, which is white and with SNR of 5 dB Fig 2 10 shows the plot of error variance vs polynomial order for a sequence $w(t_k)$, which is colored and with SNR of 5 dB Fig 2 11 shows the plot of error variance vs polynomial order for a sequence $w(t_k)$, which is colored and with SNR of 40 dB

2.5 OBSERVATIONS FROM SIMULATION -

(i) For a given SNR, the nature of error variance vs polynomial order plot remained same for given $r(t_k)$ and for different $w(t_k)$ sequences of same statistical properties, as expected The nature and location of the oscillations in the error variance vs polynomial order plot remained approximately same

(ii) The order of approximation, the nature of the nonuniform sampling grid and the nature of $w(t_k)$, are the only factors that affect the nature of $e(t_k)$

Note that the final order of approximation is decided by the nature of $r(t_k)$ as well Thus, we can say that $r(t_k)$ affects the nature of $e(t_k)$ indirectly

2.6 REASON FOR THE PRESENCE OF THE APPROXIMATING ERROR -

From the above discussion, it is clear that the orthogonal polynomial approximation procedure, seperates the high-frequency components, from the low-frequency components, present in the nonuniformly sampled signal The frequency at which this seperation occurs is decided by the polynomial order used For a more general case, where the nonuniformly sampled signal is corrupted with colored noise, if the frequency spectra of the colored nosie and the actual signal, overlap in such a way that the colored noise spectrum goes beyond the signal spectrum, then all those low-frequency components contained in

the colored noise sequence, will appear in the approximating signal In case, these spectra are nonoverlaping, then the error in the approximating signal is negligible

2.7 THE IDEA OF PREPROCESSING -

A very important observation should be mentioned at this stage The method provided in this thesis was found to give accurate estimates even for SNR levels as low as 7 dB This high accuracy can be attributed to the fact that the nonuniformly sampled data was *preprocessed* (filtered) This, effectively reduced the noise present in the data, which was subsequently used by the parameter estimator Note that these preprocessed data samples have high SNR as compared to the SNR of the nonuniformly sampled signal As a special case, the technique presented in this thesis can also be applied for uniformly sampled signals having very low SNR, and very accurate parameter estimates can be found

2.8 REMARKS -

From the discussion in this chapter it is evident that the presence of error in the approximating signal, viz , e(t), explains the inaccuracy in estimating the poles as shown by Sirkar and Sarkar [1], for considerable amount of noise Thus, if we can characterise this colored noise, we can use this additional information to obtain more accurate estimate of poles, even when the noise level is quite high

CHAPTER 3
CHARACTERIZATION OF THE APPROXIMATING ERROR

3.1 *INTRODUCTION* –

In the previous chapter we explained how a *closed-form* expression for a signal can be found by *orthogonal polynomial approximation* We also looked at the reason for the presence of the approximating error, and tried to illustrate that by simulation

In this chapter we will move one step further by trying to find out the characterization of the approximating error, and the use of this information in *parameter estimation*

It should be noted that all the simulation was done with white noise assumption for $w(t_k)$ We will show in this chapter that the white noise assumption is not stringent and the derivations given in the subsequent sections are indeed generalizations for the colored noise sequence also

We further elucidate that if the so called *normalized covariance function* of the noise, embedded in the nonuniformly sampled signal, is known a priori, then the *normalized covariance function* of the error, in the approximating signal, can be estimated

3.2 *MATHEMATICAL CHARACTERIZATION OF THE APPROXIMATING ERROR* –

We rewrite equations (2 3 1) and (2 3 11), respectively

$$x(t) = \sum_{j=1}^{N} d_j \, p_{j-1}(t) \qquad (2\ 3\ 1)$$

$$D = (P^T P)^{-1} P^T Y \qquad (2\ 3\ 11)$$

Note that, P^TP is a *diagonal matrix*, as is evident from equation (2 3 9) This is because of the fact that the approximating polynomials, p_j, are orthogonal to each other (equation 2 3 10) Furthermore, $(P^TP)^{-1}$ will also be a *diagonal matrix* Thus, the computation of $(P^TP)^{-1}$ is simplified beoause of this *orthogonality property*

Following conclusions can be drawn from equation (2 3 11), for a given nonuniform sampling grid and polynomial order

(i) From previous chapter we know that P is dependent only on the nonuniform sampling grid and the polynomial order used

(ii) Thus, $(P^TP)^{-1}P^T$ will be a constant N by K matrix

(iii) Also, the coefficient vector D is solely decided by the input vector Y

Substituting the vector D from equation (2 3 11) in equation (2 3 1) we can see that we are effectively transforming the stochastic sequence y(t) to another stochastic sequence x(t)

Let nT be the instants where x(t) is sampled uniformly (T is the inter sample distance) Therefore, equation (2 3 1) can be written as,

$$x[nT] = \sum_{j=1}^{N} d_j \, p_{j-1}[nT] \qquad (3\ 2\ 1)$$

For a given T, we can omit the notation of *nT* and replace it by *n* only We will adhere to this notation unless otherwise specified Thus, equation (3 2 1) can be expressed as,

$$x[n] = \sum_{j=1}^{N} d_j \, p_{j-1}[n] \qquad (3\ 2\ 2)$$

Rewriting equation (3 2 2) in vector form, we have,

$$X = P'D \qquad (3\ 2\ 3)$$

where X is given by,

$$X = [\ x[0]\ x[1]\ x[2] \qquad x[L-1]\ ]^T \qquad (3\ 2\ 4)$$

and L is the total number of uniformly sampled points of x(t)

hematical simplicity, we assume that,

$t_0 = 0$

$t_0 \leq nT < t_K$,

< is the total number of nonuniformly sampled points of y(t)

L by N matrix given by,

$$\begin{bmatrix} p_0[0] & p_1[0] & p_2[0] & & p_{N-1}[0] \\ p_0[1] & p_1[1] & p_2[1] & & p_{N-1}[1] \\ & & & & \\ p_0[L-1] & p_1[L-1] & p_2[L-1] & & p_{N-1}[L-1] \end{bmatrix} \qquad (3\,2\,5)$$

, given by equation (2 3 7)

ituting the value of D from equation (2 3 11) in equation (3 2 3), we get,

$$X = P' (P^T P)^{-1} P^T Y \qquad (3\,2\,6)$$

write equation (2 3 5) for Y

$$Y = [\ y(t_1)\ \ y(t_2)\ \ \ \ y(t_K)\ ]^T \qquad (2\,3\,5)$$

It is evident from equation (2 3 5) and equation (3 2 6) that, we essentially form the nonuniformly sampled input vector, Y, to a uniformly sampled output or, X

We know that $(P^T P)^{-1} P^T$ is constant for a given nonuniform sampling grid polynomial order Thus, $P' (P^T P)^{-1} P^T$ will also be constant Using this fact, we see from equation (3 2 6) that the random vector Y is transformed to another om vector X by multiplying Y with a constant L by K matrix T, given by,

$$T = P' (P^T P)^{-1} P^T \qquad (3\,2\,7)$$

Thus, each of the random variables x[0], x[1], , x[L-1] will be a linear ination of Y, given by the transformation,

$$x[n] = \sum_{m=0}^{K-1} a_{nm}\, y(t_{m+1}) \quad , \quad m = k - 1 \qquad (3\,2\,8)$$

e the new variable, m, is introduced to avoid the notation of $a_{n,\,k-1}$

Note that n corresponds to the $(n + 1)$th random variable at the instant nT

The constant *transforming coefficients* a_{nm} are such that, a_{ij} is the element ith row and jth column of the L by K transformation matrix T

From equation (2 3 9), (3 2 5) and (3 2 6), we can write the expression for a_{nm}

Appendix A)

$$a_{nm} = \sum_{i=1}^{N} \left(\frac{p_{i-1}[n]\, p_{i-1}[t_m]}{\sum_{j=1}^{K} p_{i-1}^{2}[t_j]} \right), \qquad (3\,2\,9)$$

where $p_{i-1}[n] \equiv p_{i-1}[nT]$

We know that the polynomials p_j are *deterministic* Thus, $p_j(t_n)$ will be elated with $p_j(t_{n+k})$

PROBABILITY DISTRIBUTION FUNCTION OF THE APPROXIMATING ERROR –

A *probability distribution function* (PDF) that is frequently used to model the istical behaviour of a random variable is the *Gaussian Distribution* We assume to be complex, white or colored, Gaussian, stationary stochastic process Thus, mplex random variable $g(t_k)$ at the t_kth instant, with mean $\mu_{g_{t_k}}$ and

ance $\sigma_{g_{t_k}}^{2}$, is distributed according to a Gaussian or Normal distribution if

PDF is given by,

$$p(g_{t_k}) = \frac{1}{\pi \sigma_{g_{t_k}}^{2}} \, e^{-\left(\frac{|g_{t_k} - \mu_{g_{t_k}}|^{2}}{\sigma_{g_{t_k}}^{2}} \right)} \qquad (3\,3\,1)$$

e the complex random variable, g_{t_k}, is given by,

$$g_{t_k} = u_{t_k} + j\, v_{t_k}$$

I u_{t_k} and v_{t_k} are *independent real* random variables, distributed as,

$$u_{t_k} \sim N\left(\mu_{u_{t_k}}, \frac{\sigma_{g_{t_k}}^2}{2} \right)$$

$$v_{t_k} \sim N\left(\mu_{v_{t_k}}, \frac{\sigma_{g_{t_k}}^2}{2} \right)$$

so,

$$\mu_{g_{t_k}} = E\left(g_{t_k} \right) = E\left(u_{t_k} \right) + j\, E\left(v_{t_k} \right)$$

$$= \mu_{u_{t_k}} + j\, \mu_{v_{t_k}}$$

nd,

$$\mathrm{var}(g_{t_k}) = E\left(\left| (u_{t_k} + j\, v_{t_k}) - (\mu_{u_{t_k}} + j\, \mu_{v_{t_k}}) \right|^2 \right)$$

$$= E\left(\left| u_{t_k} - \mu_{u_{t_k}} \right|^2 \right) + E\left(\left| v_{t_k} - \mu_{v_{t_k}} \right|^2 \right)$$

$$= \sigma_{g_{t_k}}^2$$

he joint PDF of the vector G, where

$G = [\ g(t_1)\ \ g(t_2)\ \ \ldots\ \ g(t_K)\ ]^T$, is given by,

$$p(G) = \frac{1}{(\pi)^K \det\left(C_{gg} \right)} e^{\left(- (G - \mu_G)^H C_{gg}^{-1} (G - \mu_G) \right)}, \qquad (3.3.2)$$

here μ_G is the mean defined as,

$$\mu_G = E\{ G \} \qquad (3.3.3)$$

nd $(\,)^H$ is the operator for the *conjugate transpose* of a matrix

is the *expectation operator*, and C_{gg} is the covariance matrix of the stochastic equence G, defined as,

$$C_{gg} = E\left([G - \mu_G] [G - \mu_G]^H \right) \qquad (3.3.4)$$

Note that μ_g is a scalar whereas μ_G is a vector

We assume that the random sequence has zero mean Therefore,

$$\mu_G = 0, \text{ null vector,} \tag{3 3 5}$$

and

$$C_{gg} = E\left(G\, G^H \right) \tag{3 3 6}$$

Substituting equation (3 3 5) and (3 3 6) in equation (3 3 2), we get,

$$p(G) = \frac{1}{(\pi)^K \det\left(C_{gg} \right)}\, e^{\left(- \left(G^H C_{gg}^{-1} G \right) \right)} \tag{3 3 7}$$

Note that C_{gg} will be a *symmetric Toeplitz matrix*, since, we have assumed $w(t_k)$ to be a *stationary process*

We now consider the effect of the polynomial approximation procedure on $w(t_k)$ Consider the case when $r(t_k)$ is a zero valued sequence Thus, $y(t_k)$ will be equal to $w(t_k)$, as can be seen from (2 2 3)
Therefore,

$$y(t_k) = 0 + w(t_k) \tag{3 3 8}$$

Also, from equation (2 3 14) we can write,

$$x[n] = 0 + e[n] \tag{3 3 9}$$

Substituting equation (3 3 8) into equation (3 2 8) we get,

$$x[n] = \sum_{m=0}^{K-1} a_{nm}\, w(t_{m+1}) \tag{3 3 10}$$

From equation (3 3 9) and equation (3 3 10), we get,

$$e[n] = \sum_{m=0}^{K-1} a_{nm}\, w(t_{m+1}) \tag{3 3 11}$$

From equation (3 3 11) we can see that the random variables e[n] are not independent of each other This is because of the fact that the polynomials p_j are deterministic Thus, $p_j(t_n)$ will be correlated with $p_j(t_{n+k})$ Note that the *orthogonality property* does not come into picture, because we are considering only one polynomial, p_j, at a time Thus, even in the case when $w(t_k)$'s are *uncorrelated*, the approximating error sequence e[n] will be *correlated* This way, the white noise assumption for $w(t_k)$, then, becomes a special case

In equation (3 3 11), we have used the same polynomial order K, as was used when $r(t_k)$ is a nonzero sequence Thus, if $w(t_k)$ are known, then e[n], which is the approximating error, can be found accurately But this statement is eluding in the sense that, if $w(t_k)$ was known then we would have subtracted it from $y(t_k)$ right at the start Thus, it becomes necessary to find the statistical properties of e[n] for parameter estimation We will now try to investigate this possibility

Equation (3 3 7) gives the joint PDF of the random sequence $w(t_k)$ We can find a similar expression for PDF of e[n], as given by (3 3 11)

Note that the joint PDF of e[n] will also be Gaussian, because, from equation (3 3 11) we can see that e[n] is a linear combination of the random variables $w(t_k)$, and each of the random variables $w(t_k)$ is assumed to have the same Gaussian PDF [4]

Thus, the PDF of e[n] can be written as,

$$p(E) = \frac{1}{(\pi)^L \det(C_{ee})} e^{\left(- (E^H C_{ee}^{-1} E) \right)} \qquad (3\ 3\ 12)$$

where,

$$E = [\ e[0]\ \ e[1]\ \ \ \ \ \ e[L-1]\ ]^T \qquad (3\ 3\ 13)$$

and

$$C_{ee} = E \left(E\ E^H \right) \qquad (3\ 3\ 15)$$

3.4 PDF OF THE APPROXIMATING ERROR FOR NON-GAUSSIAN NOISE -

We will make use of the *Central Limit Theorem* to find the PDF of the approximating error, when $w(t_k)$ has a non-Gaussian PDF

The Central Limit Theorem states that, for the random variables u_i, under general conditions, the density p(u) of their sum,

$$u = u_1 + u_2 + \qquad + u_n$$

properly normalized, tends to a normal curve as $n \to \infty$

In other words, if n is sufficiently large, then

$$p(u) = \frac{1}{\pi \sigma_u^2} e^{-\left(\frac{|u - \mu_u|^2}{\sigma_u^2} \right)}$$

It has been shown [4] that the theorem does not hold if only a small number, *m*, of the given densities are dominant, i e , if the densities of the other random variables are relatively narrow in the sense that, in the evaluation of the convolution, they can be approximated by impulses Furthermore, it has been stated [4] that p(u) is effectively the convolution of only the *m* dominant densities and it need not be close to a normal curve However, if the densities $p(u_i)$ are smooth, then p(u) is nearly Gaussian even when n is small

In our problem, the value which n takes is decided by the polynomial order N used for approximation Furthermore, this order depends upon the amount of variations in the input signal

From the simulation results given in the subsequent chapters it is shown that the Gaussian assumption for e[n] indeed holds true, for all practical purposes

3.5 THE NORMALIZED COVARIANCE / CORRELATION MATRIX -

Following are some of the properties of the covariance matrix

(i) If the stochastic sequence is stationary and white, then the covariance matrix is a diagonal matrix with equal elements The diagonal elements are equal to the variance of the process Note that, as the process is assumed to be stationary, each of the random variables will have equal variance

(ii) If the stochastic sequence is stationary and colored, then it's covariance matrix is a symmetric Toeplitz matrix

(iii) For a white nonstationary stochastic sequence, the covariance matrix is a diagonal matrix but with unequal elements

(iv) For a colored nonstationary stochastic sequence, the covariance matrix does not posses any of the special properties mentioned above

We define a nonstationary process as the one which is not a Wide Sense Stationary (WSS) [4] process A process is called WSS if its mean is constant and its covariance depends only on the lag parameter and not on the exact time instant

We now define the so called *Normalized Covariance Matrix* Note that the Normalized Covariance Matrix can be defined only for a WSS process

The Normalized Covariane Matrix is that matrix which is equal the matrix formed by normalizing all the elements of the Covariance Matrix with respect to any of its principal diagonal elements , all the principal diagonal elements of the Covariance Matrix will be equal, for a stationary process In other words, we get the Normalized Covariance Matrix by dividing each of the elements of the Covariance Matrix by the variance (for a zero mean process) Note that the covariance marix reduces to a correlation matrix for a zero mean process

The normalized covarianoe matrix is an identity matrix for a white noise stationary sequence It will be a *symmetric Toeplitz matrix*, for a colored noise stationary sequence

We can express equations (3 3 7) and (3 3 12), respectively, in terms of their normalized covariance matrices and variances

Therefore,

$$p(G) = \frac{1}{(\pi\sigma_g^2)^K \det(C'_{gg})} e^{\left(-\frac{1}{\sigma_g^2}(G^H C'^{-1}_{gg} G)\right)} \quad (3.5.1)$$

and,

$$p(E) = \frac{1}{(\pi\sigma_e^2)^L \det(C'_{ee})} e^{\left(-\frac{1}{\sigma_e^2}(E^H C'^{-1}_{ee} E)\right)} \quad (3.5.2)$$

where, C'_{gg} and C'_{ee} are the normalized covariance matrices for $g(t_k)$ and $e[n]$ respectively

In Section 3.3 we stressed on the need to find the statistical properties of $e[n]$. We will now elucidate how this can be done

Because of zero mean assumption, the covariance function is equal to the correlation function. The correlation matrix R_{gg} for $w(t)$ can be written as,

$$R_{gg} = \begin{bmatrix} r_{gg}[0] & r_{gg}[-1] & & r_{gg}[-K] \\ r_{gg}[1] & r_{gg}[0] & & r_{gg}[-(K-1)] \\ & & & \\ r_{gg}[K] & r_{gg}[K-1] & & r_{gg}[0] \end{bmatrix} \quad (3.5.3)$$

where K is the number of uniformly sampled points

But as $g(t)$ is WSS, equation (3.5.3) can be written as

$$R_{gg} = \begin{bmatrix} r_{gg}[0] & r_{gg}[1] & & r_{gg}[K] \\ r_{gg}[1] & r_{gg}[0] & & r_{gg}[K-1] \\ & & & \\ r_{gg}[K] & r_{gg}[K-1] & & r_{gg}[0] \end{bmatrix} \quad (3.5.4)$$

since,

$$r_{gg}^*[k] = r_{gg}[-k], \quad (3.5.5)$$

where, * denotes the conjugation operator

Similarly, the correlation matrix R_{ee} can be written as,

$$R_{ee} = \begin{bmatrix} r_{ee}[0] & r_{ee}[1] & \cdots & r_{ee}[L] \\ r_{ee}[1] & r_{ee}[0] & \cdots & r_{ee}[L-1] \\ \vdots & \vdots & \ddots & \vdots \\ r_{ee}[L] & r_{ee}[L-1] & \cdots & r_{ee}[0] \end{bmatrix} \quad (3.5.6)$$

Note that, both R_{gg} and R_{ee} are symmetric Toeplitz matrices, and $r_{gg}[k]$ and $r_{ee}[k]$ are the respective correlation functions, at lag k

We now rewrite equations (3.3.7), (3.3.12), (3.5.1), (3.5.2), by substituting the correlation matrix R, instead of the covariance matrix C
Therefore, we get,

$$p(G) = \frac{1}{(\pi)^K \det(R_{gg})} e^{\left(-(G^H R_{gg}^{-1} G)\right)}, \quad (3.5.7)$$

$$p(E) = \frac{1}{(\pi)^L \det(R_{ee})} e^{\left(-(E^H R_{ee}^{-1} E)\right)}, \quad (3.5.8)$$

$$p(G) = \frac{1}{(\pi\sigma_g^2)^K \det(R'_{gg})} e^{\left(-\frac{1}{\sigma_g^2}(G^H R'^{-1}_{gg} G)\right)}, \quad (3.5.9)$$

$$p(E) = \frac{1}{(\pi\sigma_e^2)^L \det(R'_{ee})} e^{\left(-\frac{1}{\sigma_e^2}(E^H R'^{-1}_{ee} E)\right)} \quad (3.5.10)$$

Note that R'_{gg} and R'_{ee} are the normalized correlation matrices of $g(t_k)$ and e[n] respectively Also,

$\sigma_g^2 = r_{gg}[0]$ and $\sigma_e^2 = r_{ee}[0]$

Following relations hold for the WSS processes g[n] and e[n] respectively

The mean is given by,

$$\mu_g = E\left(g[n] \right), \tag{3 5 11}$$

$$\mu_e = E\left(e[n] \right) \tag{3 5 12}$$

The *autocorrelation functions* (ACF) are given by,

$$r_{gg}[k] = E\left(g^*[n+k]\, g[n] \right) \tag{3 5 13}$$

$$r_{ee}[k] = E\left(e^*[n+k]\, e[n] \right) \tag{3 5 14}$$

The *crosscorrelation function* (CCF) is given by,

$$r_{ge}[k] = E\left(g^*[n+k]\, e[n] \right) \tag{3 5 15}$$

For a linear shift invariant (LSI) system with impulse response h[n] and with a WSS random input, various relationships between the input and output correlation functions and h[n] can be given

Let g[n] be the input sequence and e[n] be the output sequence
Therefore,

$$r_{ge}[k] = h[k] \circledast r_{gg}[k]$$

$$= \sum_{l=-\infty}^{\infty} h[l]\, r_{gg}[k-l] \tag{3 5 16}$$

$$r_{eg}[k] = h^*[-k] \circledast r_{gg}[k]$$

$$= \sum_{l=-\infty}^{\infty} h^*[-l]\, r_{gg}[k-l] \tag{3 5 17}$$

$$r_{ee}[k] = h[k] \otimes r_{eg}[k]$$

$$= h[k] \otimes h^{*}[-k] \otimes r_{gg}[k]$$

$$= \sum_{m=-\infty}^{\infty} h[k-m] \sum_{l=-\infty}^{\infty} h^{*}[-l]\, r_{gg}[m-l] \qquad (3.5.18)$$

where $\otimes$ denotes the discrete convolution operator

We now consider how to find the correlation function, $r_{ee}[k]$, of the approximating error $e[n]$

First step is to find $h^{*}[-k]$ and $h[k]$ from equation (3.2.6) Next we find $r_{ew}[k]$ from equation (3.5.17) Now we can use $r_{eg}[k]$ to find $r_{ee}[k]$ from equation (3.5.18) Once $r_{ee}[k]$ is known, it can be used to find the correlation matrix R_{ee} from equation (3.5.6)

Note that the computaion of R_{ee} as given above will necessitate us to know the exact value of $r_{ww}[k]$ in the first place This may not be possible in many cases This problem is solved by using the normalized correlation function instead of the actual correlation function of $w[n]$ We denote this normalized correlation function of $w[n]$ by $r'_{ww}[k]$, defined by,

$$r'_{gg}[k] = \frac{r_{gg}[k]}{r_{gg}[0]} \qquad (3.5.19)$$

Similarly we define the normalized correlation function, $r'_{ee}[k]$, of $e[n]$, by,

$$r'_{ee}[k] = \frac{r_{ee}[k]}{r_{ee}[0]} \qquad (3.5.20)$$

Note that this so called normalized correlation function preserves all the statistical properties of any WSS process

The normalized correlation function can also be interpreted as the correlation function of the same stochastic process which has unit variance

Furthermore, it is evident that if $w[n]$ is white then its normalized correlation function will be such that,

$r'[0] = 1$, and

$r'[k] = 0$, for all $k \neq 0$

and the corresponding normalised correlation matrix R' will be an identity matrix

Similar interpretation can also be made for any colored WSS process

From the above discussion we observe that for a WSS process, r[0] is proportional to the energy content in the process Furthermore, any change in the energy content will change the variance of the process leaving the normalised correlation function unaffected

In subsequent chapters we will develop methods which will use the normalized correlation function estimate, as given by equation (3 5 20), to get accurate parameter estimates

CHAPTER 4
PARAMETER ESTIMATION IN COLORED GAUSSIAN NOISE

4.1 *INTRODUCTION* -

In the previous chapters we studied the transformation from the nonuniform sampling grid domain, to the uniform sampling grid domain We also saw how colored noise is introduced because of the polynomial approximation procedure We also proved that, if the signal and noise frequency spectra do not overlap, then this colored noise is negligible The problem of finding the statistical properties of the colored noise in the approximating signal was discussed fully in the previous chapter

The purpose of this chapter will be to use the statistical properties of the colored noise during parameter estimation The signal is modelled by a sum of *complex exponentials*, embedded in colored noise whose normalized covariance matrix is known All the simulation was done on real data, which essentially becomes a special case

Lot of pioneering work in parameter estimation for complex exponentials has been done by Roy and Kailath [6], Kumaresan and Tufts [9], and Farrier et al [10] The technique discussed here considers the more general case of colored noise Furthermore, this technique exploits the properties of the noise corrupted signal *autocorrelation-like* matrix by using *subspace separation approach*

It is well known that a signal consisting of complex exponentials can be modelled by a *difference equation* We have presented a technique which uses this difference equation, and achieves accurate estimates of complex exponentials using *Singular Value Decomposition* [11]

Further enhancement in accuracy is achieved by using the signal subspace seperation approach, which is also referred as *Principal Components Analysis* [12]

4.2 MATHEMATICAL CHARACTERIZATION OF THE APPROXIMATING SIGNAL -

The approximating signal, x(t), consisting of M complex exponentials, can be expressed as,

$$x(t) = \sum_{i=1}^{M} A_i e^{s_i t} + e(t) \qquad (4.2.1)$$

Let x[0], x[1], , x[L-1] be the *uniformly sampled* corrupted sequence Thus, the *discrete-time* process, x[n], can be modelled by a time-series as,

$$x[n] = r[n] + e[n] \qquad (4.2.2)$$

where, r[n] is the nth sample of the noncorrupted signal term, given by,

$$r[n] = - \sum_{k=1}^{M} a[k]\, r[n-k]$$

which is the *difference equation* representation of the *Prony's model* [1] The coefficients a[k] are such that the impulse response of the filter, A(z), given by,

$$A(z) = \frac{1}{1 + a[1]z^{-1} + a[2]z^{-2} + \quad + a[M]z^{-M}} \qquad (4.2.3)$$

is equal to $\sum_{k=1}^{M} A_i e^{s_i n}$, for $n \geq M$

Let s_i and z_i be the roots in the s and z domain, respectively The transformation from the continuous to discrete-time domain is achieved through,

$$z_i = e^{s_i T} \qquad (4.2.4)$$

where, T is the inter sample time period, and s_i is given by,

$$s_i = \alpha_i + j\omega_i = \alpha_i + j2\pi f_i \qquad (4.2.5)$$

where, α_i is the damping factor and f_i is the *normalized frequency*

The error, e[n], is an innate part of the model and gives rise to the random nature of the observed process x[n] In our case, this is the colored approximating error sequence

From equation (4 2 2), it follows that,

$$\begin{aligned} x[M] &= -a[1]\, r[M-1] - \cdots - a[M]\, r[0] + e[M] \\ x[M+1] &= -a[1]\, r[M] - \cdots - a[M]\, r[1] + e[M+1] \end{aligned} \qquad (4\ 2\ 6)$$

$$x[L-1] = -a[1]r[L-1] - \cdots - a[M]r[L-M-1] + e[M-1]$$

Equation (4 2 6) can be written in vector notation as,

$$X = QA + E \qquad (4\ 2\ 7)$$

where,

$$X = [\ x[M]\ \ x[M+1] \quad \cdots \quad x[L-1]\ ]^T \qquad (4\ 2\ 8)$$

$$Q = \begin{bmatrix} r[M-1] & \cdots & r[0] \\ r[M] & \cdots & r[1] \\ \vdots & & \vdots \\ r[L-2] & \cdots & r[L-M-1] \end{bmatrix} \qquad (4\ 2\ 9)$$

$$A = [\ -a[1]\ \ -a[2] \quad \cdots \quad -a[M]\ ]^T \qquad (4\ 2\ 10)$$

and

$$E = [\ e[M]\ \ e[M+1] \quad \cdots \quad e[L-1]\ ]^T \qquad (4\ 2\ 11)$$

The joint PDF of E, as given by equation (4 2 11), can be written as,

$$p(E) = \frac{1}{(\pi\sigma_e^2)^L \det(R'_{ee})}\, e^{-\left(\frac{1}{\sigma_e^2}\left(E^H R'^{-1}_{ee} E\right)\right)} \qquad (4\ 2\ 12)$$

where, the normalized correlation matrix, R'_{ee}, is found by the method mentioned in the previous chapter. Substituting equation (4 2 7) in equation (4 2 12), we can write,

$$p(X, A) = \frac{1}{(\pi\sigma_e^2)^L \det(R'_{ee})}\, e^{-\left(\frac{1}{\sigma_e^2}\left(X - QA\right)^H R'^{-1}_{ee}\left(X - QA\right)\right)} \qquad (4\ 2\ 13)$$

where, the notation p(X, A) explicitly shows the dependence of the joint probability density function on both X and A. Note that X, Q and R'_{ee} are known quantities, while A and σ_e^2 are unknowns.

We find the *maximum likelihood estimate* (MLE) of A by maximizing equation (4.2.13) with respect to A [12]. This can be done equivalently, by minimizing,

$$S = \left(X - QA \right)^H R'^{-1}_{ee} \left(X - QA \right) \qquad (4.2.14)$$

which is a quadratic function of A. After minimizing equation (4.2.14) with respect to A, we get the estimate of A as,

$$A = \left(Q^H R'^{-1}_{ee} Q \right)^{-1} Q^H R'^{-1}_{ee} X \qquad (4.2.15)$$

and the value of the quadratic function, S, is,

$$S_{MIN} = X^H \left(R'^{-1}_{ee} - R'^{-1}_{ee} Q \left(Q^H R'^{-1}_{ee} Q \right)^{-1} Q^H R'^{-1}_{ee} \right) X \qquad (4.2.16)$$

We rewrite equation (4.2.15) as,

$$\left(Q^H R'^{-1}_{ee} Q \right) A = Q^H R'^{-1}_{ee} X \qquad (4.2.17)$$

Therefore,

$$\Gamma A = B \qquad (4.2.18)$$

where,

$$\Gamma = \left(Q^H R'^{-1}_{ee} Q \right) \qquad (4.2.19)$$

and

$$B = Q^H R'^{-1}_{ee} X \qquad (4.2.20)$$

Thus, Γ can be interpreted as the estimate of the signal autocorrelation-like matrix, where the signal is corrupted by colored noise.

We can write the noise corrupted signal *autocorrelation-like matrix*, Γ, as, (See Appendix B)

$$\Gamma = R_{ss} + R_{ee} \qquad (4.2.21)$$

We perform spectral decomposition of R_{ss} and R'_{ee} Thus, we get,

$$\Gamma = \sum_{i=1}^{M} \lambda_i^s u_i u_i^H + \sigma_e^2 \sum_{i=1}^{M} \lambda_i^e u_i u_i^H \qquad (4\,2\,22)$$

where, λ_i^s is the ith eigen value of the signal autocorrelation-like matrix and λ_i^e is the ith eigen value of the colored noise normalised autocorrelation matrix, and both R_{ee} and R_{ss} are of dimension M by M, M being the number of complex exponentials present in the corrupted signal Note that both the signal and noise subspaces are spanned by the same eigen vectors, u_i, of Γ This is because an M-dimensional space can be spanned by any set of M orthogonal vector in the space We observe that, as the signal is locally WSS [4], the column and row eigen vectors of R_{ss} are same (See Appendix C) Same argument is also applicable for the autocorrelation matrix of colored noise

In almost all practical applications, the value of M is unknown Thus, we assume some large value, p, possibly much larger than M
We rewrite equation (4 2 22) as,

$$\Gamma = \sum_{i=1}^{p} \lambda_i^s u_i u_i^H + \sigma_e^2 \sum_{j=1}^{p} \lambda_i^e u_i u_i^H \qquad (4\,2\,23)$$

As R_{ss} is of dimension p by p and it's rank is only M (M << p), it is not *full rank* The sample autocorrelation-like matrix, Γ, is, however, full rank due to the inclusion of the $\sigma_e^2 \lambda_i^e$ term In the special case, when e[n] is white, λ_i^e are unity

The frequeny and damping factor information about the signal is contained in the eigenvectors u_i Since, Γ, as given by equation (4 2 19), is not the actual signal autocorrelation matrix but an estimate of it, the estimate, $\hat{\Gamma}$, can be expressed as,

$$\hat{\Gamma} = \sum_{i=1}^{M} \left(p\,\hat{\lambda}_i^s + \hat{\sigma}_e^2 \hat{\lambda}_i^e \right) \hat{v}_i \hat{v}_i^H + \hat{\sigma}_e^2 \sum_{j=M+1}^{p} \hat{\lambda}_i^e \hat{v}_i \hat{v}_i^H \qquad (4\,2\,24)$$

This is evident from the fact that,

$\lambda_i^s = 0$, for $i = M+1, \ldots, p$

It should be noted that, Γ, as defined by equation (4 2 19), is not the exact value of the noise corrupted signal autocorrelation matrix To avoid the discrepancy to propogate any further, we have used $p\hat{\lambda}_i^s$ (instead of $\hat{\lambda}_i^s$) in the

principal eigen values in equation (4 2 24) Furthermore, the eigen vectors, $\hat{0}_1$, are related to the signal eigen vectors, e_1, by the equation,

$$\hat{0}_1 \simeq \frac{1}{(p)^{1/2}} e_1 \qquad (4\,2\,25)$$

4 3 SINGULAR VALUE DECOMPOSITION -

We will use SVD for three reasons

(i) To form the *spectral decomposition* as given by equation (4 2 24), and use it to find the estimate, $\hat{A}$

(ii) To find the model order M, from the *spectral decomposition*

(iii) To seperate the *signal* and *noise subspaces* and thereby achieve further accuracy in parameter estimates

By applying SVD [11], we decompose a n by m matrix A, such that,

$$A = U \sum V^H \qquad (4\,3\,1)$$

where,

$$U = [\ u_1\ u_2 \quad u_l\ u_{l+1} \quad u_n\ ] \qquad (4\,3\,2)$$

$$V = [\ v_1\ v_2 \quad v_l\ v_{l+1} \quad v_n\ ] \qquad (4\,3\,3)$$

and

$$\sum = \mathrm{diag}(\ \$_1\ \$_2 \quad \$_{l-1}\ \$_l\ 0\ 0 \quad 0\ ) \qquad (4\,3\,4)$$

where, $\$_1 \geq \$_2 \geq \geq \$_l > 0$,

and l is the rank of the matrix A

Also, U and V are *unitary matrices*, i e , $U^{-1} = U^H$ and $V^{-1} = V^H$, respectively

The *column* and *row eigen vectors*, u_1 and v_1, respectively, are defined as,

$$A\,u = \$\,v \qquad (4\,3\,5)$$

and,

$$A^H v = \$\,u \qquad (4\,3\,6)$$

Multiplying equation (4 3 5) by A^H and using equation (4 3 6), we get,

$$A^H A u = \zeta A^H v = \zeta^2 u \qquad (4\ 3\ 7)$$

Thus, u is the eigen vector of $A^H A$, with ζ^2 as the corresponding eigen value
Multiplying equation (4 3 6) by A and using equation (4 3 5), we get,

$$A A^H v = \zeta A u = \zeta^2 v \qquad (4\ 3\ 8)$$

Thus, v is the eigen vector of $A A^H$, with ζ^2 as the corresponding eigen value

If A is a square matrix, then one can find its *eigen value decomposition* (EVD) We will see how SVD can be used to find the EVD

From the basic properties of EVD, we know that A (A^H) and $A^H A$ (AA^H) have the same eigen vectors Also, if λ is the eigen value corresponding to an eigen vector of A (or A^H), then the eigen value of $A^H A$ (or AA^H) is equal to λ^2 Thus, if A is a square matrix, then its eigen value is the square root of the corresponding singular value [11]

For the noise corrupted signal autocorrelation-like matrix, $\hat{\Gamma}$ (which is a square p by p matrix), we can use SVD to find the spectral decomposition as given by equation (4 2 24) Note that $\hat{\Gamma}$ is found using equation (4 2 19) Also, as $\hat{\Gamma}$ is symmetric, the row and column eigen vectors will be same

Substituting the value of $\hat{\Gamma}$ from equation (4 2 24) in equation (4 2 17), we get,

$$\left(\sum_{i=1}^{M} \left(p \hat{\lambda}_i^s + \hat{\sigma}_e^2 \hat{\lambda}_i^e \right) \Phi_i \Phi_i^H + \hat{\sigma}_e^2 \sum_{i=M+1}^{p} \hat{\lambda}_i^e \Phi_i \Phi_i^H \right) \hat{A} = Q^H R'^{-1}_{ee} X$$

(4 3 9)

Equation (4 3 9) can be written in matrix notation as,

$$\begin{bmatrix} \hat{\Diamond}_1 & \hat{\Diamond}_2 & \cdots & \hat{\Diamond}_M & \cdots & \hat{\Diamond}_p \end{bmatrix} \begin{bmatrix} p\,\hat{\lambda}_1^{s} + \hat{\sigma}_e^{2}\,\hat{\lambda}_1^{e} & & & & & \\ & p\,\hat{\lambda}_2^{s} + \hat{\sigma}_e^{2}\,\hat{\lambda}_2^{e} & & & & \\ & & \ddots & & & \\ & & & p\,\hat{\lambda}_M^{s} + \hat{\sigma}_e^{2}\,\hat{\lambda}_M^{e} & & \\ & & & & \hat{\sigma}_e^{2}\,\hat{\lambda}_{M+1}^{e} & \\ & & & & & \hat{\sigma}_e^{2}\,\hat{\lambda}_p^{e} \end{bmatrix} \begin{bmatrix} \hat{\Diamond}_1^{H} \\ \hat{\Diamond}_2^{H} \\ \vdots \\ \hat{\Diamond}_M^{H} \\ \vdots \\ \hat{\Diamond}_p^{H} \end{bmatrix} \hat{A}$$

$$= Q^{H} R'^{-1}_{ee} Q \qquad (4\,3\,10)$$

Therefore,

$$\begin{bmatrix} \hat{\Diamond} & \sum & \hat{\Diamond}^{H} \end{bmatrix} \hat{A} = Q^{H} R'^{-1}_{ee} X \qquad (4\,3\,11)$$

where,

$$\hat{\Diamond} = \begin{bmatrix} \hat{\Diamond}_1 & \hat{\Diamond}_2 & \cdots & \hat{\Diamond}_M & \cdots & \hat{\Diamond}_p \end{bmatrix} \qquad (4\,3\,12)$$

and

$$\sum = \mathrm{diag}\left(\left(p\,\hat{\lambda}_1^{s} + \hat{\sigma}_e^{2}\,\hat{\lambda}_1^{e}\right) \cdots \left(p\,\hat{\lambda}_M^{s} + \hat{\sigma}_e^{2}\,\hat{\lambda}_M^{e}\right) \left(\hat{\sigma}_e^{2}\,\hat{\lambda}_{M+1}^{e}\right) \cdots \left(\hat{\sigma}_e^{2}\,\hat{\lambda}_p^{e}\right) \right)$$

As $\hat{\Diamond}$ is a unitary matrix, $\hat{\Diamond}^{-1} = \hat{\Diamond}^{H}$

Therefore, equation (4 3 11) becomes,

$$\hat{A} = \left(\hat{\Diamond} \sum^{-1} \hat{\Diamond}^{H} \right) Q^{H} R'^{-1}_{ee} X \qquad (4\,3\,13)$$

$$\hat{A} = \left(\sum_{i=1}^{M} \frac{1}{\hat{\lambda}_i^{s} + \hat{\sigma}_e^{2}\,\hat{\lambda}_i^{e}} \hat{\Diamond}_i \hat{\Diamond}_i^{H} + \sum_{i=M+1}^{p} \frac{1}{\hat{\sigma}_e^{2}\,\hat{\lambda}_i^{e}} \hat{\Diamond}_i \hat{\Diamond}_i^{H} \right) Q^{H} R'^{-1}_{ee} X$$

(4 3 14)

Therefore,

$$\hat{A} = \left(\sum_{i=1}^{p} \frac{1}{\lambda_i} \Diamond_i \Diamond_i^H \right) Q^H R'^{-1}_{ee} X \qquad (4.3.15)$$

where λ_i are the eigen values found from the EVD of $\hat{\Gamma}$

Example -

$$r[n] = 1.5\, e^{-\left(\frac{1}{75} + j\,2\pi\,0.08\right) nT} + 1.5\, e^{-\left(\frac{1}{90} + j\,2\pi\,0.11\right) nT} + 1.5\, e^{-\left(\frac{1}{75} - j\,2\pi\,0.08\right) nT} + 1.5\, e^{-\left(\frac{1}{90} - j\,2\pi\,0.11\right) nT},$$

$K = 50,\ L = 100$

Fig 4.1 shows the pole estimates using equation 4.3.15, for 10 different trials. The white noise corrupted signal is sampled at 50 nonuniformly spaced time instants. As the data consists of two sinusoids, we have two complex conjugate pairs of signal poles in the z and s domains. The estimate of the normalised autocorrelation matrix, R'_{ee}, is calculated using equations 3.5.17, 3.5.18 and 2.5.19. The SNR in this case is 40 dB and the value of the model order, p, is 10. Fig 4.2 and Fig 4.3 show the pole estimates for SNR of 20 dB and 10 dB, respectively, with p = 10. (Mention about the actual pole locations)

Figures 4.4, 4.5 and 4.6, show the pole estimates for SNRs of 40 dB, 20 dB and 10 dB, with an increased value of p, equal to 16.

The table given below shows the performance of the estimator for different noise levels and with varying values of the selected model order, p. The estimation procedure was repeated for 10 trials, thus, the bias and variance, mentioned below, will give a rough understanding about the statical performance of the estimator.

Table 4.1 - p = 10

	SNR = 40 dB		SNR = 20 dB		SNR = 10 dB	
	Bias	Variance	Bias	Variance	Bias	Variance
α_1	6.234×10^{-4}	9.959×10^{-7}	9.203×10^{-4}	1.159×10^{-7}	1.651×10^{-3}	4.073×10^{-5}
f_1	8.037×10^{-4}	5.126×10^{-8}	9.127×10^{-4}	2.771×10^{-7}	1.847×10^{-3}	4.017×10^{-6}
α_2	4.282×10^{-5}	3.421×10^{-7}	1.324×10^{-3}	6.004×10^{-6}	1.542×10^{-3}	1.574×10^{-3}
f_2	9.849×10^{-5}	7.100×10^{-8}	2.347×10^{-3}	8.303×10^{-7}	7.574×10^{-3}	1.574×10^{-3}

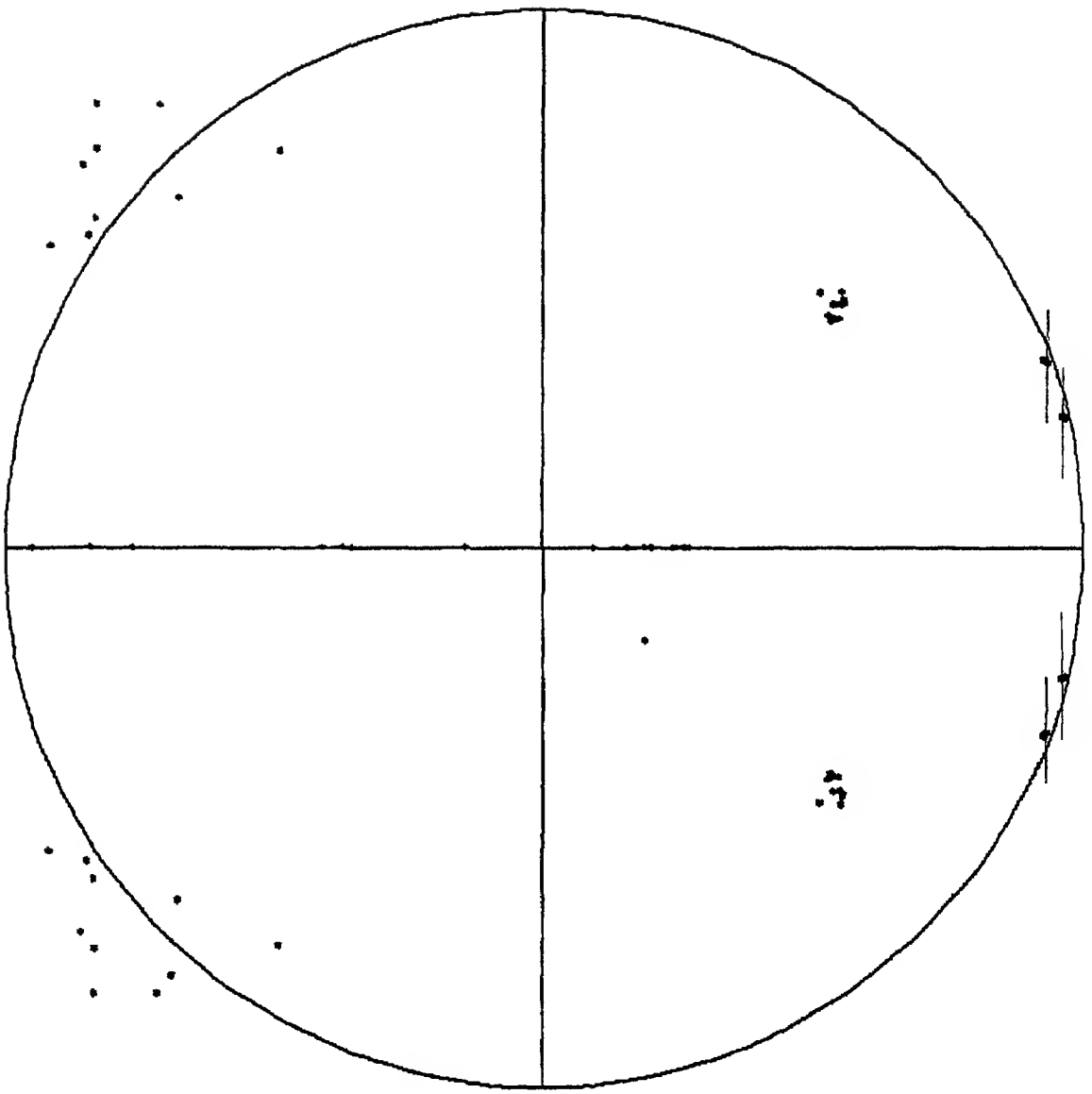

Fig 4.1

SNR = 40 dB, ρ = 10.

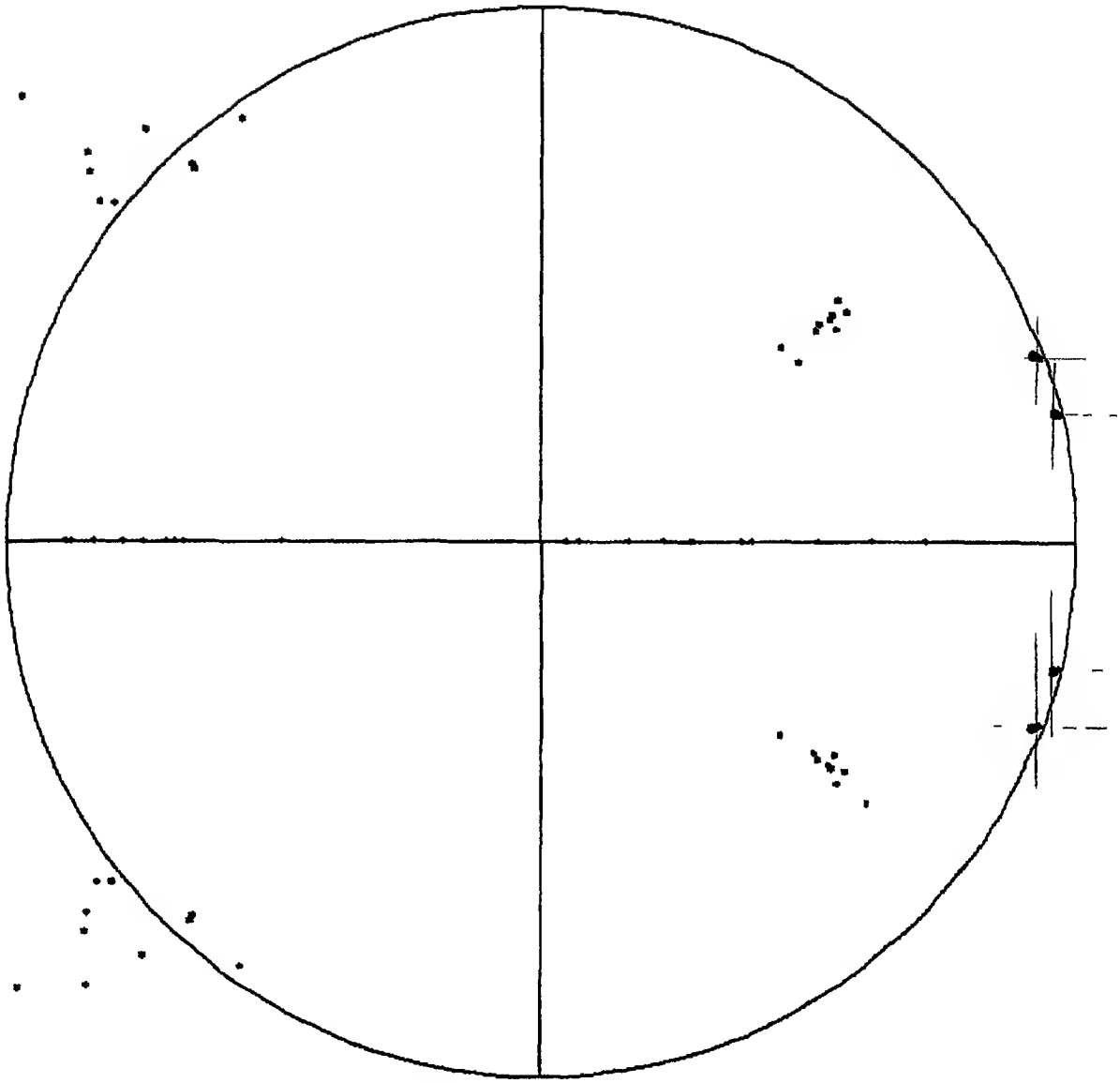

Fig 4.2

SNR = 20 dB, p = 10.

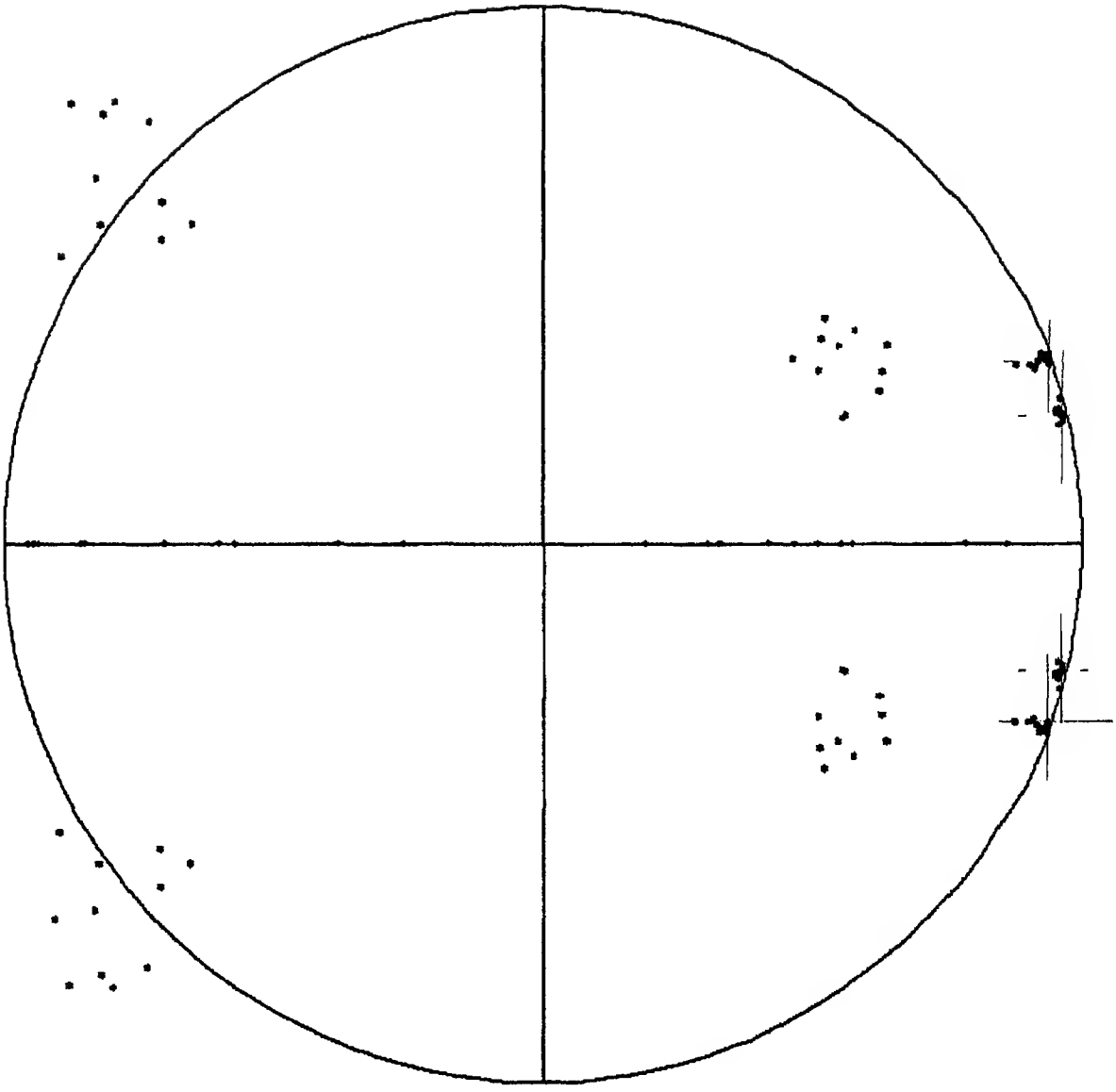

Fig 4.3

SNR = 10 dB, α = 10.

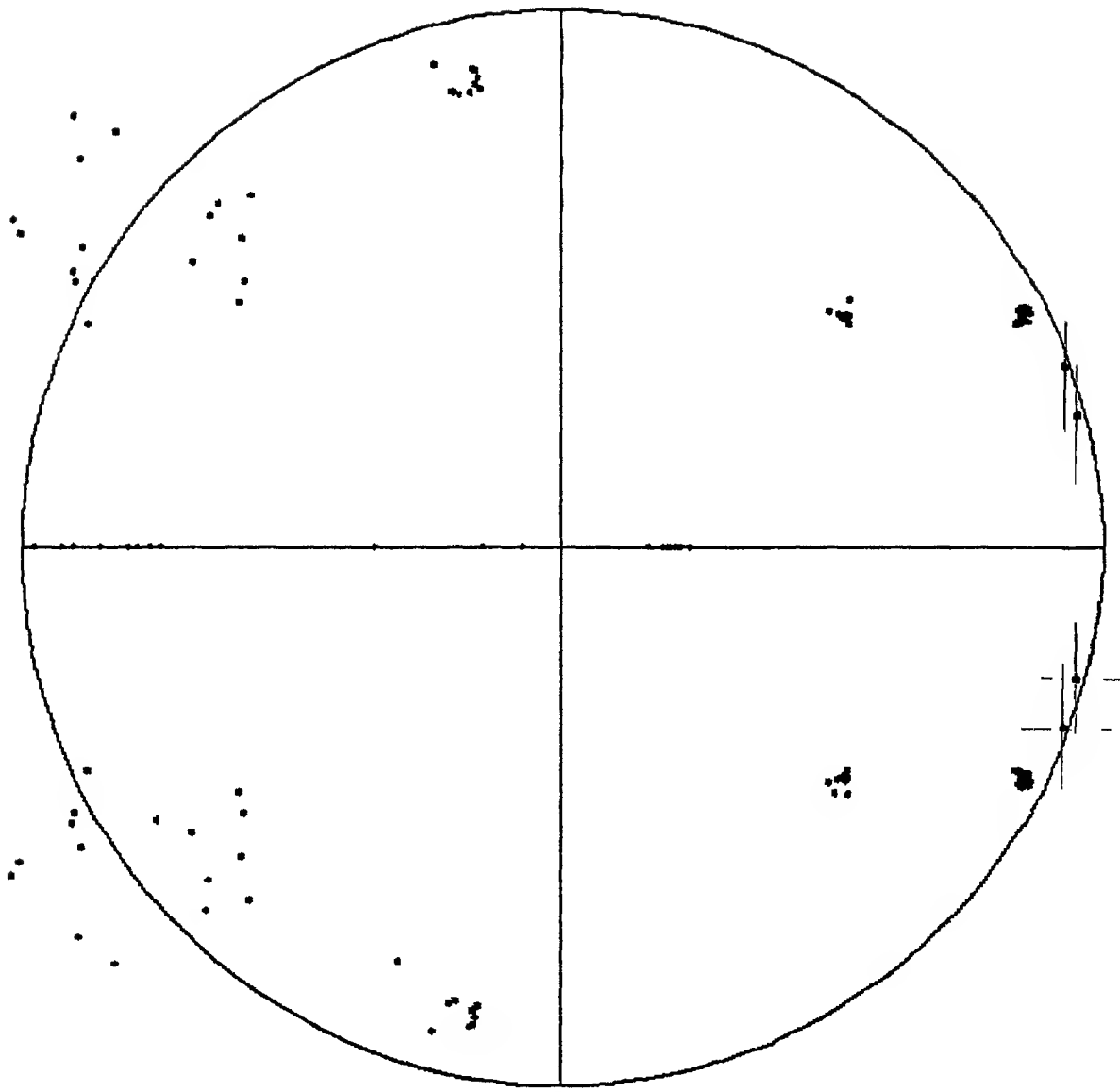

Fig 44

SNR = 40 dB, p = 16

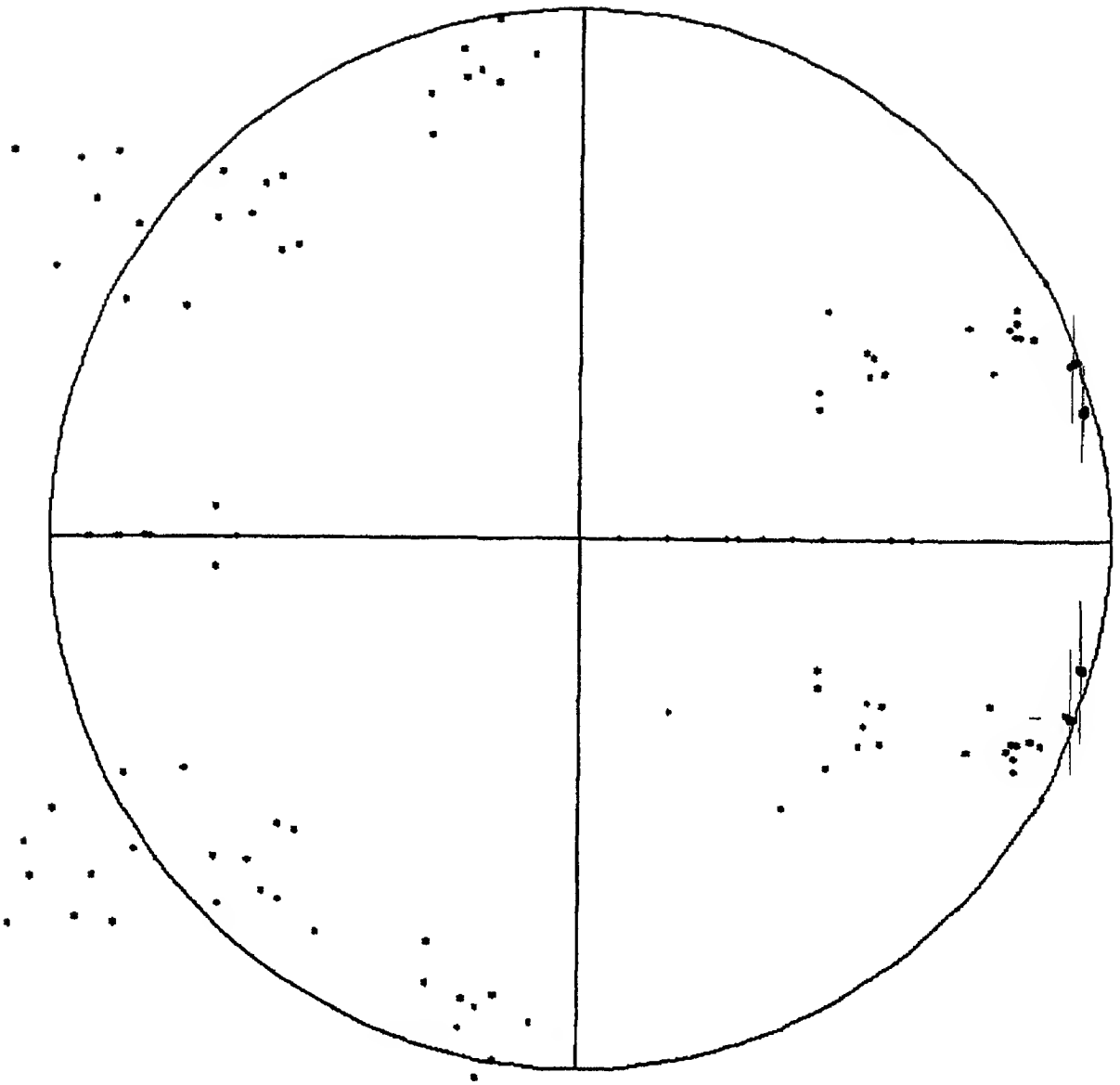

Fig. 4.5

SNR = 20 dB, p = 16.

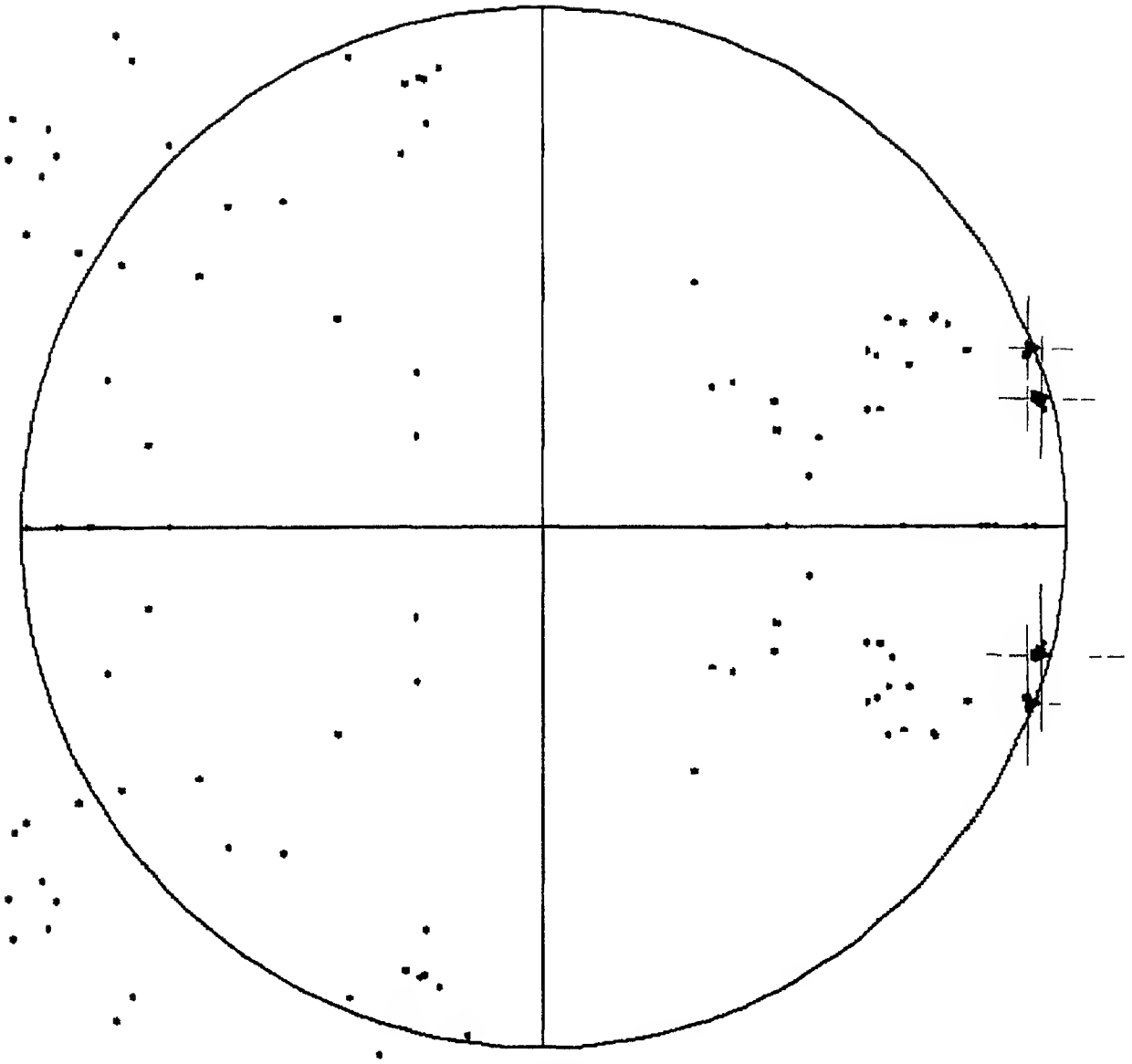

Fig 4.6

SNR = 10 dB. n = 16.

le 4.2 - p = 16

SNR = 40 dB		SNR = 20 dB		SNR = 10 dB	
Bias	Variance	Bias	Variance	Bias	Variance
$2\,576 \times 10^{-4}$	$1\,174 \times 10^{-7}$	$7\,880 \times 10^{-4}$	$8\,303 \times 10^{-6}$	$1\,418 \times 10^{-3}$	$1\,418 \times 10^{-5}$
$7\,659 \times 10^{-6}$	$2\,675 \times 10^{-8}$	$8\,240 \times 10^{-5}$	$2\,360 \times 10^{-7}$	$1\,431 \times 10^{-4}$	$2\,353 \times 10^{-6}$
$4\,118 \times 10^{-5}$	$2\,859 \times 10^{-7}$	$6\,330 \times 10^{-5}$	$2\,235 \times 10^{-6}$	$5\,521 \times 10^{-4}$	$7\,067 \times 10^{-6}$
$5\,677 \times 10^{-5}$	$5\,772 \times 10^{-9}$	$1\,492 \times 10^{-4}$	$2\,649 \times 10^{-7}$	$2\,621 \times 10^{-4}$	$1\,027 \times 10^{-6}$

Following conclusions can be drawn from the location of the pole estimates, different SNR levels and different values of p

1) The pole estimates are accurate down to the SNR level of 10 dB
1) As the SNR is decreased, the variance of estimated poles increases accordingly These perturbations are introduced because of the ill-conditioning of the problem [9] In Section 4 5 we will discuss how this ill-conditioning can be taken care of, with corresponding increase in accuracy
1) The variance of the estimated poles is reduced by increasing the model order, p We will elucidate this fact in the subsequent section

MODEL ORDER SELECTION THROUGH SPECTRAL DECOMPOSITION -

In this section we will address the problem of model order selection from a ven data set Equation (4 3 10) will be used for this purpose From equation 3 10), the *principal eigen values*, viz , ($p\hat{\lambda}_i^s + \hat{\sigma}_e^2 \hat{\lambda}_i^e$ i = 1, , M), can selected, such that they are large as compared to the *nonprincipal eigen lues*, viz , ($\hat{\sigma}_e^2 \hat{\lambda}_i^e$ i = M+1, , p), through proper choice of p Note it, as $\hat{\lambda}_i^e$ are the eigen values of the normalized autocorrelation matrix of the proximating error, they are very small in magnitude, when the noise variance, $\hat{\sigma}_e^2$, is also small Thus, the M principal eigen values are *distinctly large* than the - M nonprincipal eigen values, *for a proper choice of p* This fact can be used r appropriately selecting the model order, M

rom equation (4 2 24), it is evident that the estimate of the autocorrelation-like atrix, $\hat{\Gamma}$, can be written as,

$$\hat{\Gamma} = \hat{A} + \hat{B} \qquad (4\ 4\ 1)$$

here, $\hat{A}$ and $\hat{B}$ are defined to be the principal and nonprincipal terms, espectively

We maintain that principal and nonprincial eigenvalues are easily istinguishable This fact is due to the fundamental property of spectral ecomposition, that,

$$\min \| \hat{\Gamma} - \hat{A} \| = \hat{\lambda}_{z+1} \qquad (4\ 4\ 2)$$

or any $z < p$, where, $\hat{\lambda}_i$ are the eigen values of $\hat{\Gamma}$

In another words, a full rank matrix $\hat{\Gamma}$, whose rank is p, can be approximated y a low rank marix $\hat{A}$, whose rank is z, where $z < p$, and the approximation error is iven by equation (4 4 2)

Thus M corresponds to that minimum order after which the rate of change of he Mth eigen value, with respect to z, is almost constant, i e ,

$$\frac{d [\min \| \hat{\Gamma} - \hat{A} \|]}{dz} \rightarrow \text{constant}$$

le define this rate of change as,

$$\frac{d [\min \| \hat{\Gamma} - \hat{A} \|]}{dz} = \frac{\hat{\lambda}_{z+1} - \hat{\lambda}_{z+2}}{1} \qquad (4\ 4\ 3)$$

Thus, z is a correct order, if it corresponds to that minimum order, after hich the difference between the (z+1)th and (z+2)th eigen value is small as ompared to the difference between the zth and (z+1)th eigen values

During the process of simulation, it was observed that equation (4 4 3) is not ery usable for model order selection One reason could be that the estimate of he autocorrelation-like matrix is not exact, due to limited amount of data With ncreased amount of data, it is only expected that equation (4 4 3) may give ccurate model order estimates

An alternative approach for model order selection has been proposed by idzow and Wu [13] which examines the *norm ratio*, R The norm ratio, R, is defined

$$R = \left(\frac{\lambda_1^2 + \lambda_2^2 + \quad + \lambda_q^2}{\lambda_1^2 + \lambda_2^2 + \quad + \lambda_p^2} \right)^{\frac{1}{2}}, \qquad (4\,4\,4)$$

here, q is the actual model order, and p is the model order used for simulation, ith p > q, and $(\lambda_i \quad i = 1, 2, \quad p)$ are the corresponding eigen value estimates f the autocorrelation-like matrix We can easily see that the norm ratio R always es between zero and one Moreover, if M readily identifiable largest eigen value stimates exist, then the smallest value of q for which the rank ratio, R, tends to nity, will correspond to the actual model order, M [13]

Practically it is observed that the the ratio of the largest principal ingular value, $\$_1$, to the smallest principal singular value, $\$_M$, of the estimated utocorrelation-like matrix, is large (Note that $\lambda_i^2 = \$_i$, for i = 1, 2, , p), or any SNR level The transition from $\$_M$ to $\$_{M+1}$, is evident only for moderate NR levels, of about 20 dB Selecting the appropriate model order becomes ifficult for SNR below this threshold, using traditional methods But as we have *reprocessed* the data by using *orthogonal polynomial approximation*, the SNR of he approximating signal is increased, comparatively Due to this fact (it was erified from simulation), the transition from $\$_M$ to $\$_{M+1}$, is evident even uptill NR of 7dB, when the data is *preprocessed* Note that, in our case the polynomial pproximation procedure was necessary, to enable the transformation from onuniform sampling grid to uniform sampling grid Furthermore, we have shown hrough simulation in Chapter 5 that, by using the idea of preprocessing, for a niformly sampled data, we get very accurate pole estimates as compared to raditional methods

Keeping the practical considerations in mind, we propose a slightly ifferent criteria for selecting the appropriate model order

We define a criterion C, such that,

$$C = \frac{\$_i}{\$_{i+1}}, \text{ for } i = 1, 2, \quad p-1 \qquad (4\,4\,5)$$

Thus, the order, i, will be equal to the actual model order, M, when C is maximum

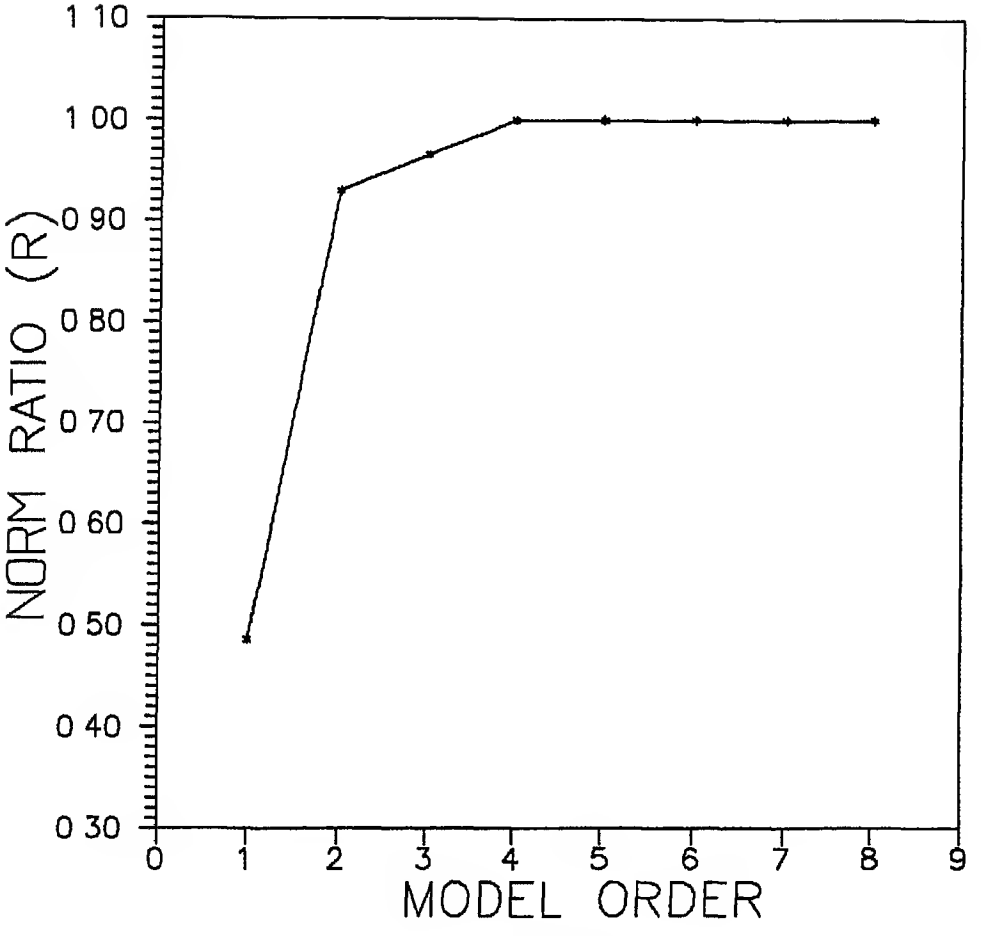
1 10
1 00
0 90
0 80
0 70
0 60
0 50
0 40
0 30
NORM RATIO (R)
0
1
2
3
4
5
6
7
8
9
MODEL ORDER

Fig 4 7

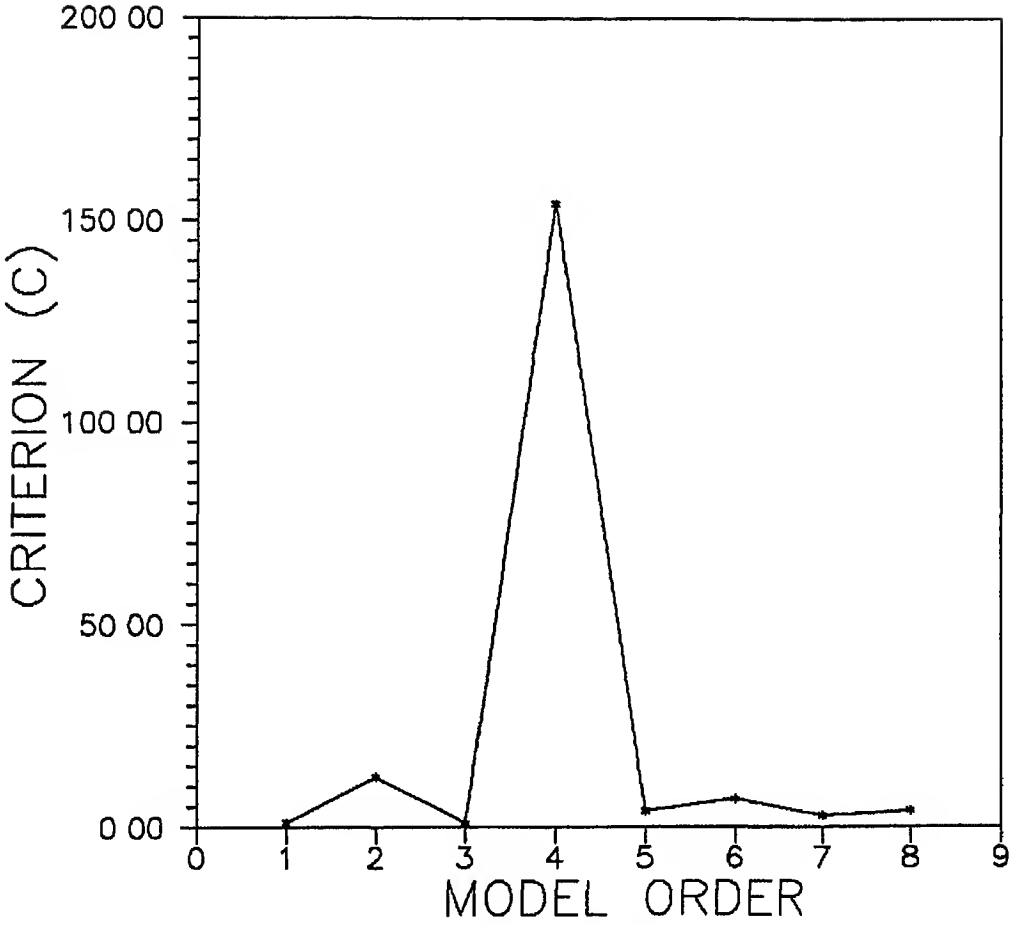
200 00
150 00
100 00
50 00
0 00
CRITERION (C)
0
1
2
3
4
5
6
7
8
9
MODEL ORDER

Fig 4.8

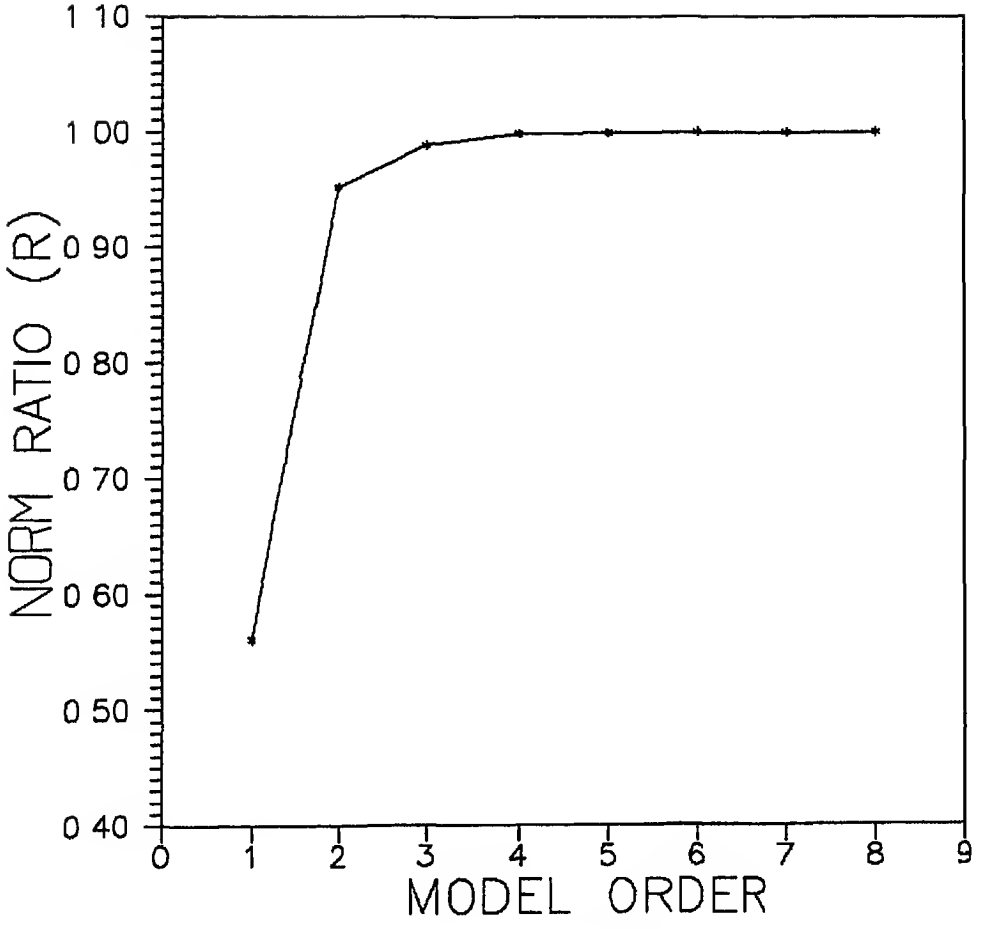
1 10
1 00
0 90
0 80
0 70
0 60
0 50
0 40
NORM RATIO (R)
0
1
2
3
4
5
6
7
8
9
MODEL ORDER

Fig 4.9

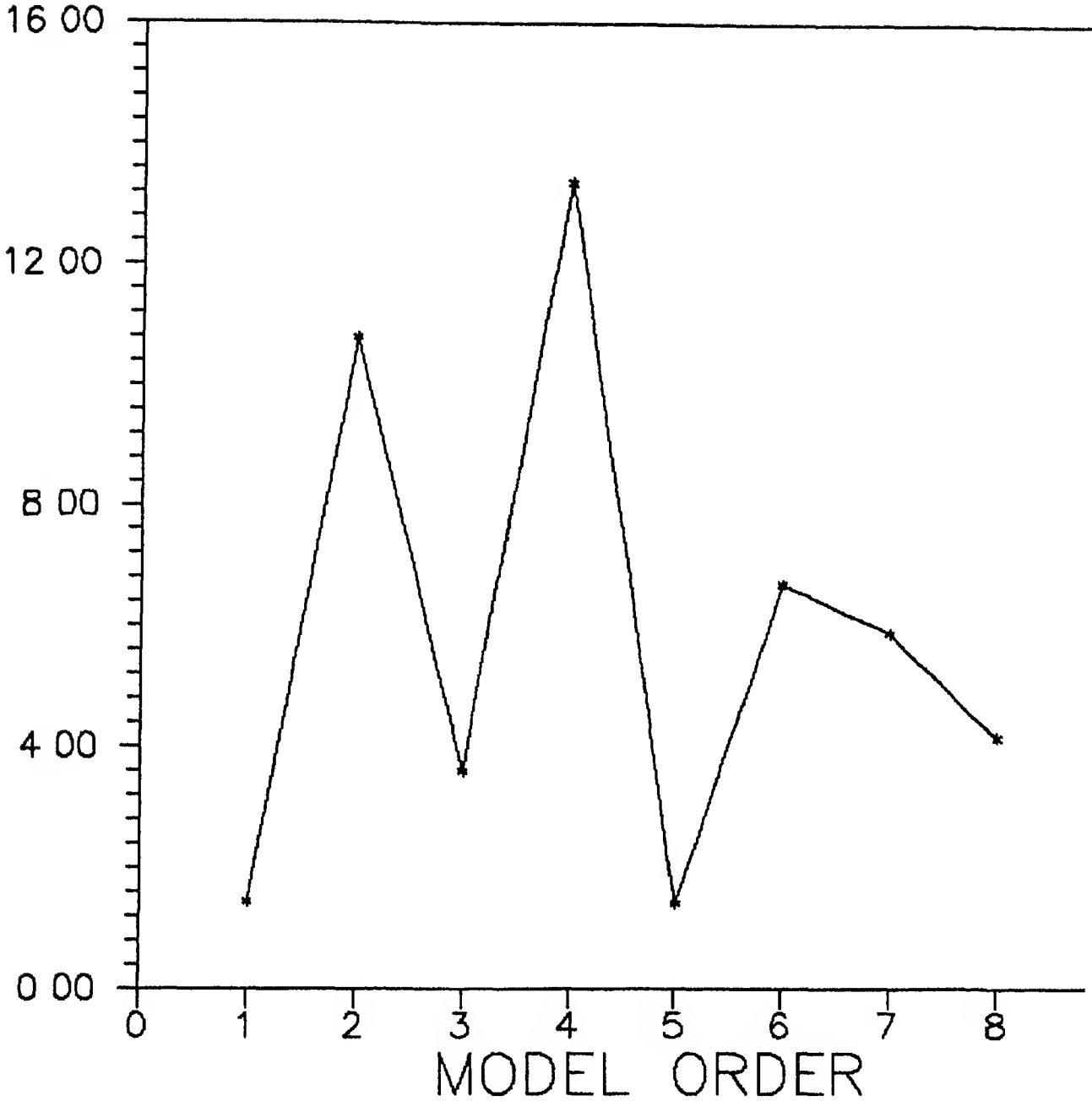

Fig410

Fig 4 7 shows a typical plot of the norm ratio, R, vs the model order, for SNR of 7dB, and with p equal to 20 (Note that the value of p is large) M was 4 for this simulation Fig 4 8 shows a typical plot of the criterion, C, vs the model order, for the same example Comparing these two cases, we observe that the actual model order, M, can be easily distinguished by using the criteria given by equation (4 4 5), for the example under consideration Fig 4 9 and Fig 10 show the plots for the R and C vs model order, respectively, but for p = 16 We see from Fig 4 9 and Fig 4 10 that it is more difficult to select the model order when p is reduced

Note that the value of p should be selected in such a manner that the smallest principal eigen value (i e , the Mth eigen value) is sufficiently large than any of the nonprincipal eigen values This is to say that, any decrease in the SNR level should be accompanied by a corresponding increase in the value of p, so that the principal eigen values are easily distinguishable If this condition is not met, then model order selection becomes an impossible task, for very low SNR levels

4.5 *THE SUBSPACE SEPARATION APPROACH* -

Upto this point, we discussed how to find the parameter estimate, $\hat{A}$, using equation (4 3 13) In the previous section we also saw how spectral decomposition helps us in selecting the model order In this section, we will formulate a technique, which achieves further accuracy in the parameter estimates, than that can be achieved using equation (4 3 13)

We will assume that model order selection has already been done The order selected will correspond to the actual order, M, only if the SNR is not below, say 7 dB This statement has been explained in the previous section

Now that M is known, we can effect the subspace seperation, as given by equation (4 4 1), where $\hat{A}$ uses only the principal eigen vectors and eigen values As M corresponds to that minimum order for which the error norm, defined by equation (4 4 2), is minimum, we can approximate $\hat{\Gamma}$, by $\hat{A}$ (which is of rank M) Thus, the new estimator can be written, such that it uses only the M principal eigen vectors and eigen values Thus,

$$\hat{A}_{PC} = \left(\sum_{i=1}^{M} \frac{1}{p\,\hat{\lambda}_i^{s} + \hat{\sigma}_e^{2}\,\hat{\lambda}_i^{e}} \hat{v}_i \hat{v}_i^{H} \right) Q^{H} R'^{-1}_{ee} X \qquad (4\,5\,1)$$

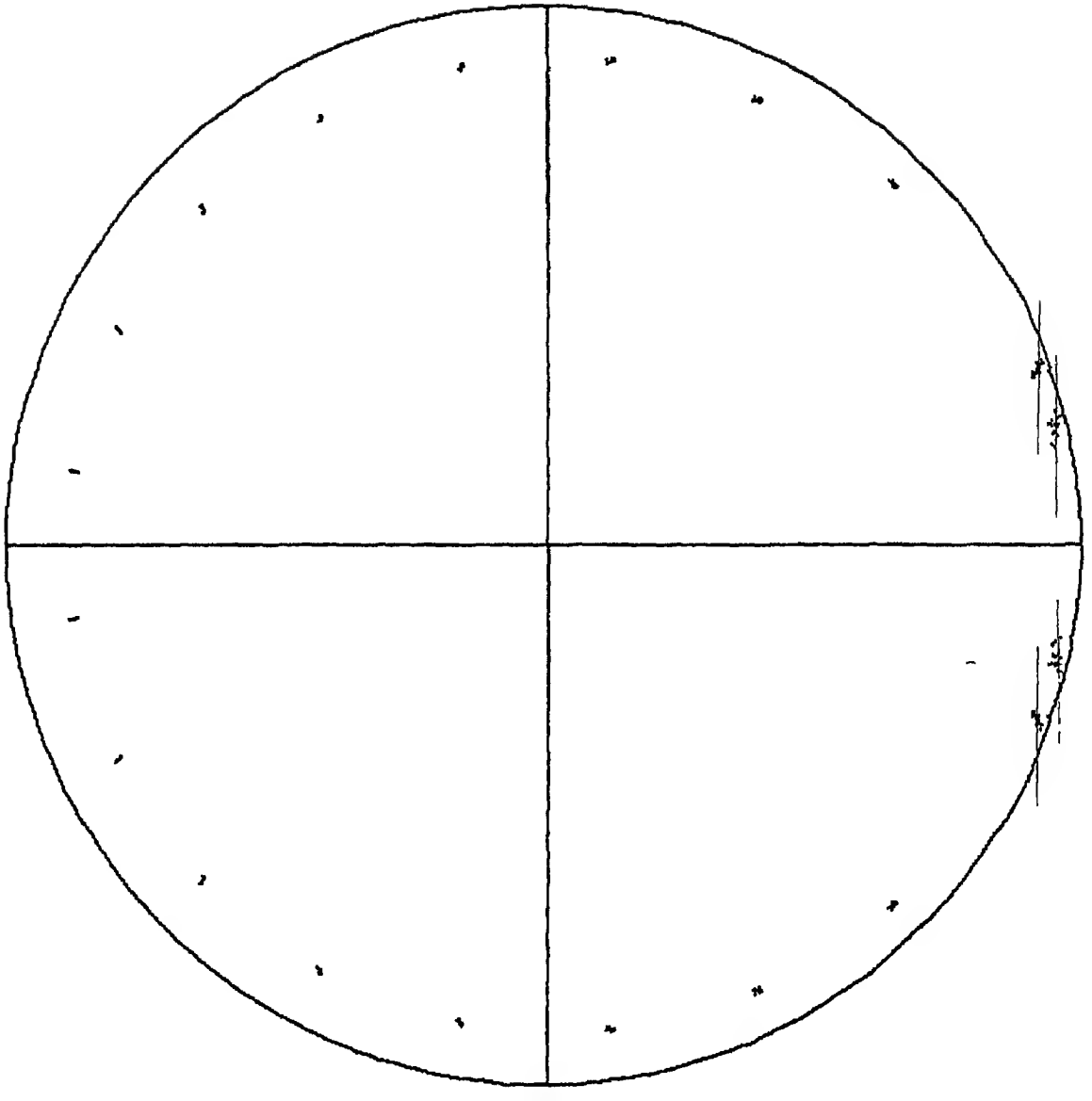

Fig. 4.11.

No compensation for radial bias.

Only principal eigen values are used.

SNR ≡ 7 dB, p ≡ 20. Plot for 20 realizations.

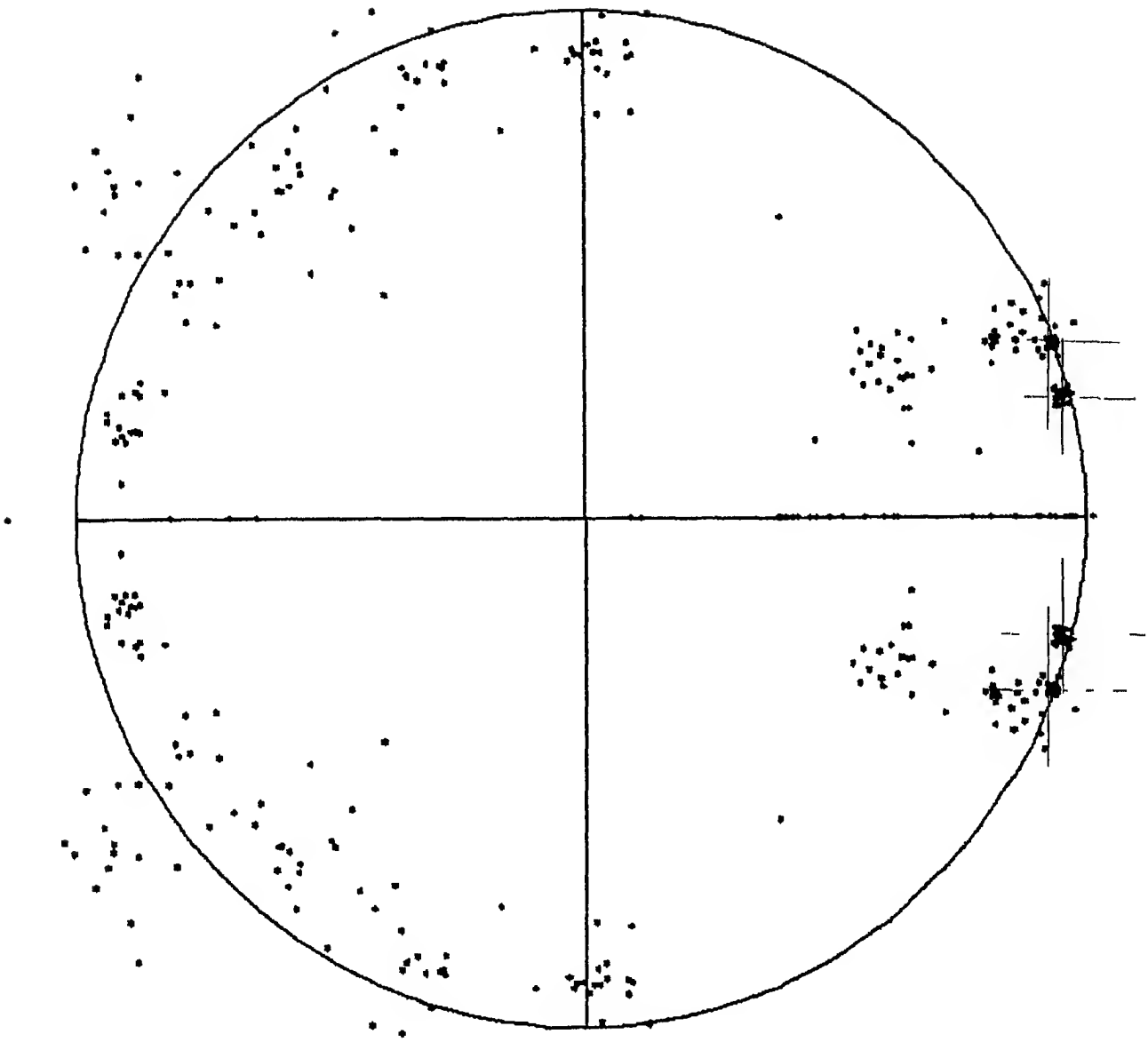

Fig 4.12

All eigen values are used.

SNR = 7 dB, p = 20 Plot for 20 realizations

A similar approach (but restricted to the white noise case) has also been used in Ref 9 The subspace separation approach mentioned here is referred as *truncated SVD* in Ref 9 The principal difference between the approach given here and the approach used in Ref 9 is that, we use an autocorrelation-like matrix associated with the time series unlike the data matrix itself as done in Ref 9 It was observed that the autocorrelation-like matrix gives more accurate estimates as compared to the estimates from the data matrix itself This is due to the fact that the noise effects are considerably moderated by using the autocorrelation-like matrix along with a large value of p

Fig 4 11 shows the pole estimates for 20 different realizations of the random sequence, $w(t_k)$, using the estimator given by equation (4 5 1), with SNR of 7 dB and p equal to 20 Fig 4 12 shows the pole estimates for 20 different realizations but using the estimator given by equation (4 3 15), for the same case Clearly, it can be seen that the variance of the pole estimates shown in Fig 4 11 has reduced substantially as compared to the variance of the pole estimates shown in Fig 4 12 We also observe a certain amount of *radially inward shift* in the location of the pole estimates in the z domain of Fig 4 11 This *phenomenon of radial bias* will be explained in detail in section 4 7

4.6 FURTHER IMPROVEMENTS IN THE SUBSPACE SEPARATION APPROACH –

We write the spectral decomposition as given by equation (4 5 1)

$$\hat{\Gamma} = \sum_{i=1}^{M} \left(p\,\hat{\lambda}_i^{s} + \hat{\sigma}_e^{2} \right) \hat{v}_i \hat{v}_i^{H} \qquad (4\,6\,1)$$

where, the estimated eigen vectors, $\hat{v}_i$, are related to the actual signal eigen vectors, e_i, by equation (4 2 25) See [14] for properties of the autocorrelation-like matrix, $\hat{\Gamma}$

The matrix $\hat{B}$, which is spanned by the p - M nonprincipal eigen vectors is given by,

$$\hat{B} = \sum_{i=M+1}^{p} \hat{\sigma}_e^{2}\, \hat{\lambda}_i^{e}\, \hat{v}_i \hat{v}_i^{H} \qquad (4\,6.2)$$

Using the estimate $R'_{ee}{}^{-1}$, we can effect it's spectral decomposition, to find the

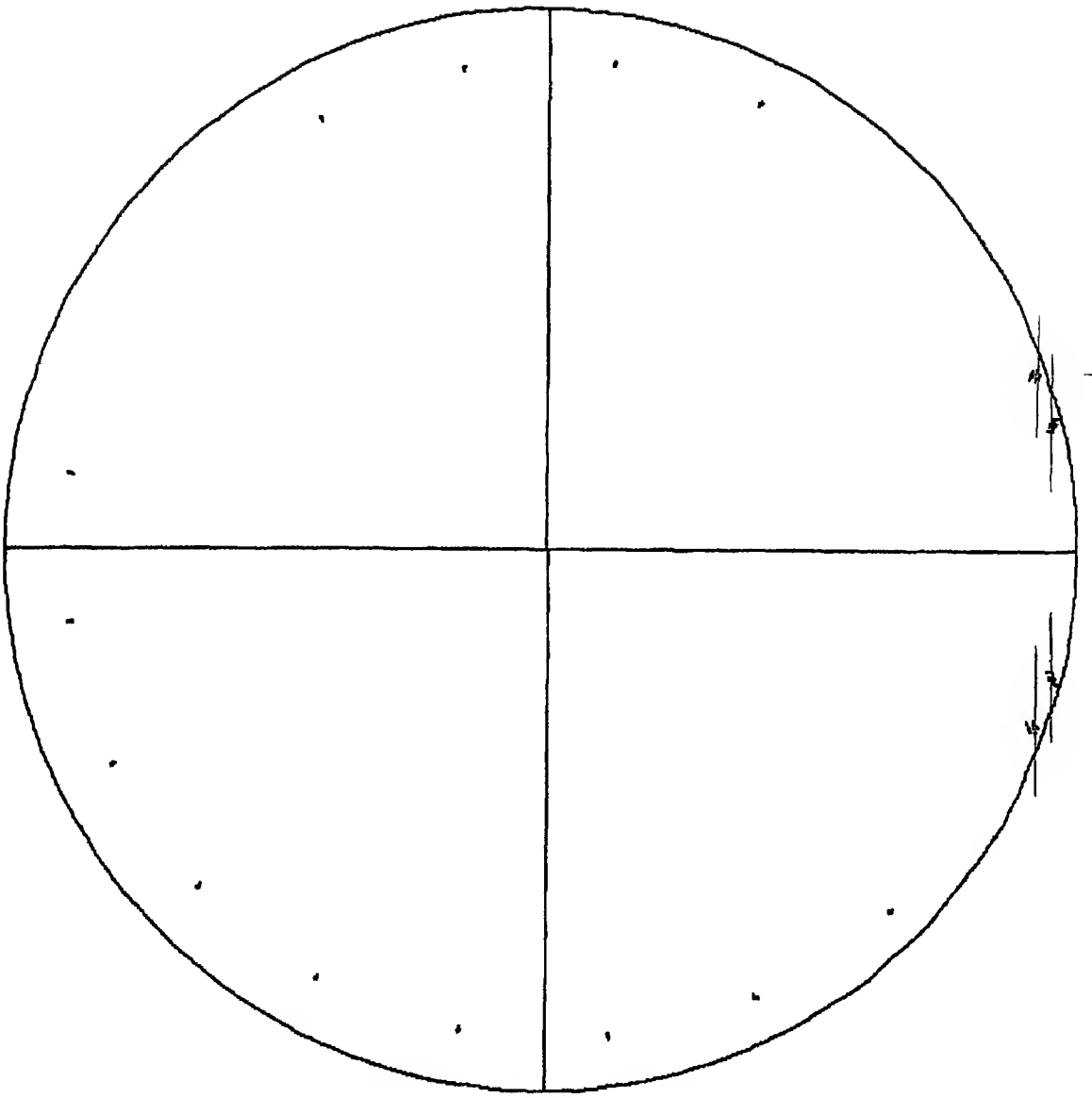

Fig 4.13

Compensation for radial bias.

Only principal eigen values are used.

SNR = 7 dB, p = 20. Plot for 20 realizations.

estimates ($\hat{\lambda}_i^e$ $i = 1, 2, \ldots, p$). Using these eigen values in equation (4.6.2), we can find the estimate of the colored noise variance, $\hat{\sigma}_e^2$. It was observed during simulation that the noise variance estimate is correct, only when p is moderately high. Thus, using the colored noise variance estimate, $\hat{\sigma}_e^2$, and ($\hat{\lambda}_i^e$ $i = 1, 2, \ldots, M$), we can subract the terms ($\hat{\sigma}_e^2$ $\hat{\lambda}_i^e$ $i = 1, 2, \ldots, M$), from the M principal eigen values of $\hat{\Gamma}$, as given by equation (4.6.1). Thus, for high values of p, it is only expected that the autocorrelation-like matrix, $\hat{\Gamma}$, can be written as,

$$\hat{\Gamma} = \sum_{i=1}^{M} p \hat{\lambda}_i^s \hat{v}_i \hat{v}_i^H \qquad (4.6.3)$$

Thus, the new relation for finding the parameter estimate, $\hat{A}_{MPC}$, is given by,

$$\hat{A}_{MPC} = \left(\sum_{i=1}^{M} \frac{1}{p \hat{\lambda}_i^s} \hat{v}_i \hat{v}_i^H \right) Q^H R'^{-1}_{ee} X \qquad (4.6.4)$$

The pole estimates, using the estimator given by equation (4.6.4), for 20 different realizations, are shown in Fig. 4.13. M is 4, p is 20 and SNR is 7 dB. Comparing Fig. 4.11 and Fig. 4.13, we observe that the *radial bias* has reduced. The actual pole locations are shown by cross in Fig. 4.9 and Fig. 4.11.

4.7 *RADIAL BIAS* -

We will explain the phenomenon of radial bias using equations (4.6.1) and (4.6.2). (In Ref. [9], the phenomenon of radial bias, is depicted, for white noise case, on page 837).

The eigen vectors, e_i and e_i^H, of the signal autocorrelation matrix, R_{ss}, can be written as,

$$e_i = \left(1 \quad e^{1(\alpha_i + j\omega_i)} \quad e^{2(\alpha_i + j\omega_i)} \quad \ldots \quad e^{(p-1)(\alpha_i + j\omega_i)} \right)^T,$$

and,

$$e_i^H = \left(1 \quad e^{1(\alpha_i - j\omega_i)} \quad e^{2(\alpha_i - j\omega_i)} \quad \ldots \quad e^{(p-1)(\alpha_i - j\omega_i)} \right)^T \qquad (4.7.1)$$

and,

$$R_{ss} = \sum_{i=1}^{M} \lambda_i e_i e_i^H \qquad (4.7.2)$$

where,

$\lambda_i = A_i^2$ = power of the ith sinusoid in the signal s[n], given by,

$$s[n] = \sum_{i=1}^{M} A_i e^{s_i n}, \qquad (4.7.3)$$

and,

$$s_i = \alpha_i + j 2\pi f_i \qquad (4.7.4)$$

From equation (4.6.1), we can see that the estimate of λ_i, say λ'_i, which we calculate, is always greater than the actual value of λ_i

Therefore,

$$\lambda'_i > \lambda_i \qquad (4.7.5)$$

Thus, for any ith sinusoid,

$$\lambda'_i u_i u_i^H = \left(\hat{\lambda}_i^s + \frac{\hat{\sigma}_e^2 \hat{\lambda}_i^e}{p} \right) u_i u_i^H$$

$$> \lambda_i u_i u_i^H \qquad (4.7.6)$$

From equation (4.7.6), we observe that, the increase in the estimated eigen values is due to the term $\frac{\hat{\sigma}_e^2 \hat{\lambda}_i^e}{p}$ Also, if e[n] is a white noise sequence, then $\hat{\lambda}_i^e$ will be unity

Thus, we can write,

$$\lambda'_i e^{s_i n} > \lambda_i e^{s_i n}$$

i.e.,

$$\lambda'_i e^{\alpha_i n} e^{j 2\pi f_i n} > \lambda_i e^{\alpha_i n} e^{j 2\pi f_i n}$$

Therefore,

$$k \lambda_i e^{\alpha_i n} e^{j 2\pi f_i n} > \lambda_i e^{\alpha_i n} e^{j 2\pi f_i n}, \text{ where } k \lambda_i = \lambda'_i$$

Therefore,

$$\lambda_i e^{\alpha'_i n} e^{j 2\pi f_i n} > \lambda_i e^{\alpha_i n} e^{j 2\pi f_i n}, \text{ where } k e^{\alpha_i n} = e^{\alpha_i n}$$

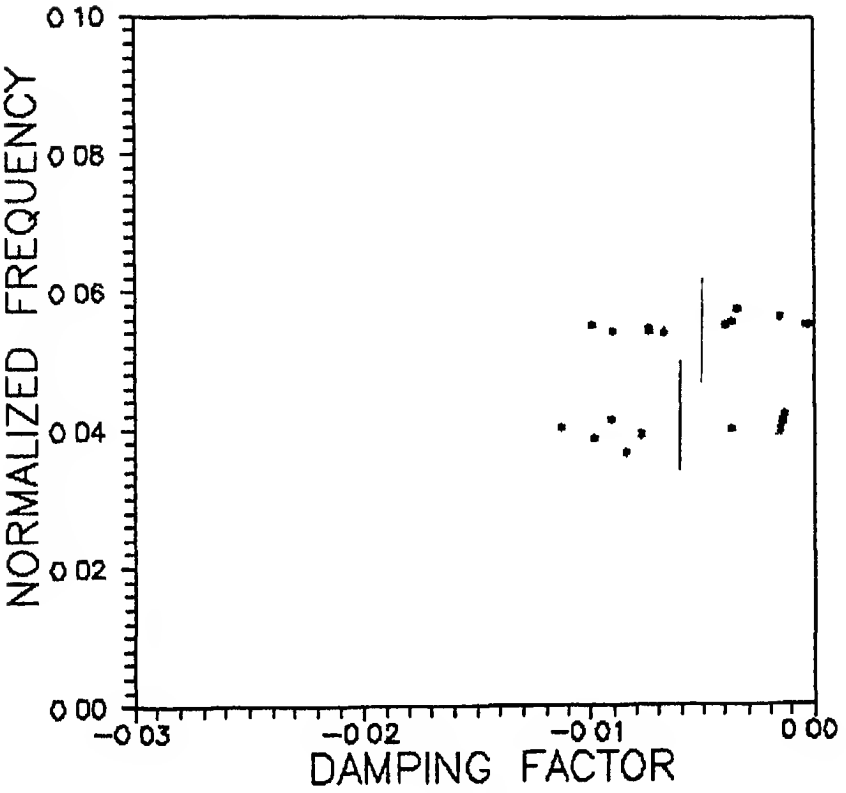

Fig 4.14

s - domain representation.

Therefore,

$$\alpha'_1 n = \alpha_1 n + \ln k \qquad (4\,7\,7)$$

Note that $\hat{\lambda}_1^e$ are always positive, since, they represent the power of the corresponding sinusoidal components Also, as $\hat{\sigma}_e^2$ and p are positive, it is evident from equation (4 7 6) that the term $\frac{\hat{\sigma}_e^2 \hat{\lambda}_1^e}{p}$ will always be positive Clearly, the multiplying factor, k, in equation (4 7 7) will also be positive Moreover, it will always be grater than or equal to unity The later condition will be applicable for no noise conditions Thus, α'_1 will always be greater than α_1, for any nonzero value of $\hat{\sigma}_e^2$

We observe that, the effect of noise is to increase the damping factor It is well know that an increase of damping factor, in s domain, shifts the corresponding pole near the center of the unit circile, in z domain, i e , the pole in the z domain moves radially inward This radial inward shift is termed as *radial bias* From equation (4 7 6) we note that, in no noise conditions, this radial shift is negligible

Furthermore, we maintain that the angular displacemant of the poles in z domain should theoretically be almost negligible This means that, idealy, the frequency estimates must be very accurate, even for low SNRs

Note that the entire left half of the s-plane is mapped in the unit circle in the z-plane Moreover, a pole at minus infinity in s-plane is equivalent to the pole at the oente of the unit circle in z-plane Thus, if the actual pole is away from the unit circle but inside it, then even a slight radial perturbation in its position in the z-plane will amount to a logarithmic change of its position in the s-plane Thus, while comparing simulation results, the position of the pole in the z-plane plays a very important role, since, the same percentage of change in the pole position, when it is away from the unit circle, amounts to a very large error The unit circle repesentation, hence, is a somewhat elusive repesentation However, in applications, where the maximum and minimum damping factors differ widely, the unit circle representation will be more convenient because of its concise nature

Fig 4 14 shows two (of the four) signal pole estimates in the s domain for the case depicted in Fig 4 6 Comparing Fig 4 14 and Fig 4 6 we observe that the variations in the damping factor estimates are more apparent from Fig 4 14 than from Fig 4 6 Thus, the s domain representation is more appropriate

CHAPTER 5
UNIFORM SAMPLING AS A SPECIAL CASE OF NONUNIFORM SAMPLING

5.1 *INTRODUCTION* -

In this chapter, we will consider the special case of parameter estimation from a uniformly sampled data The purpose of this chapter is to enlighten the noise cancelling features inherent in the orthogonal polynomial approximation procedure The effectiveness of the technique given in this thesis, will be clear from simulation results It is only expected that, same accuracy in pole estimates will be maintained for other similar cases

5.2 *EXAMPLE 1* -

In this example, we will use equation (2 2 3) to approximate the given uniform samples This is to say that, $\{t_k \quad k = 1, 2, \quad , K\}$ will be uniformly distributed (Note that we are considering uniform sampling as a special case of nonuniform sampling) Thus, the white noise corrupted samples, $y(t_k)$, will represent the given uniformly sampled corrupted signal, whose parameters are to be estimated In this example, we will consider the specific time-series, of $r(t_k)$, given by,

$$r(t_k) = A_1 e^{s_1 t_k} + A_2 e^{s_2 t_k} + A_3 e^{s_3 t_k} + A_4 e^{s_4 t_k} + w(t_k) \qquad (5\,2\,1)$$

where, $w(t_k)$ are zero mean independent, real, Gaussian stationary random variables, with unknown variance, σ_w^2, and

$s_1 = 5\,55 \times 10^{-3} + j\,0\,08, \quad s_2 = 5\,55 \times 10^{-3} - j\,0\,08,$
$s_3 = 6\,66 \times 10^{-3} + j\,0\,11, \quad s_4 = 6\,66 \times 10^{-3} - j\,0\,11,$
$A_1 = 3\,5, \quad A_2 = 3\,5, \quad A_3 = 1\,5, \quad A_4 = 1\,5,$
and, t_k are uniformly spaced

First, we *preprocess* $y(t_k)$ using orthogonal polynomial approximation, to generate the approximating signal, x[n], as given by equation (4 2 2) Note that the SNR in x[n] is greater than the SNR in $y(t_k)$ Once the uniform samples, x[n], n = 0, 1, , L, are known, we use the estimator given by equation (4 6 4), to to find

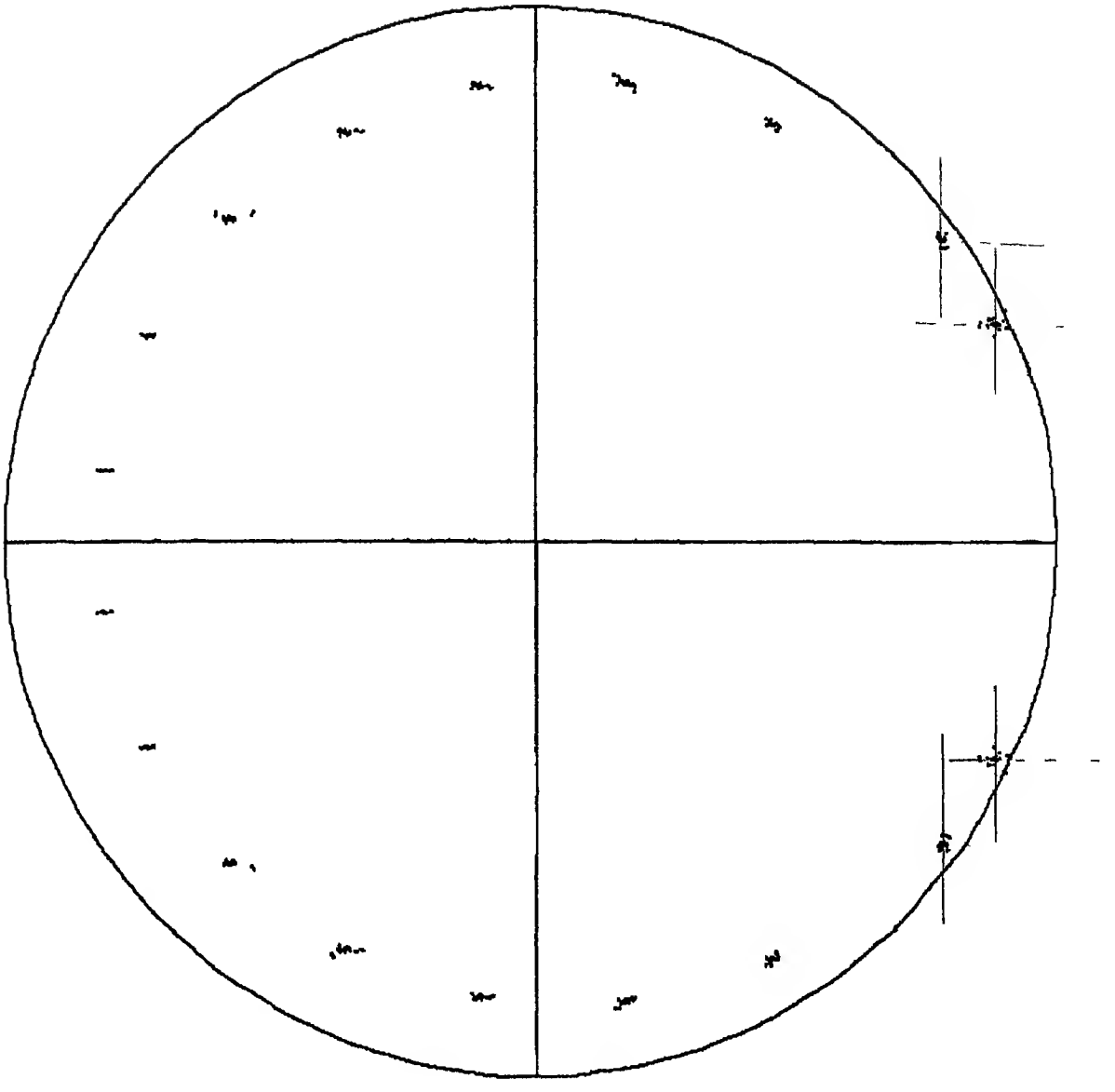

Fig. 5.1

With preprocessing.

SNR = 6 dB, p = 20.

Uniform sampling as a special case of

nonuniform sampling

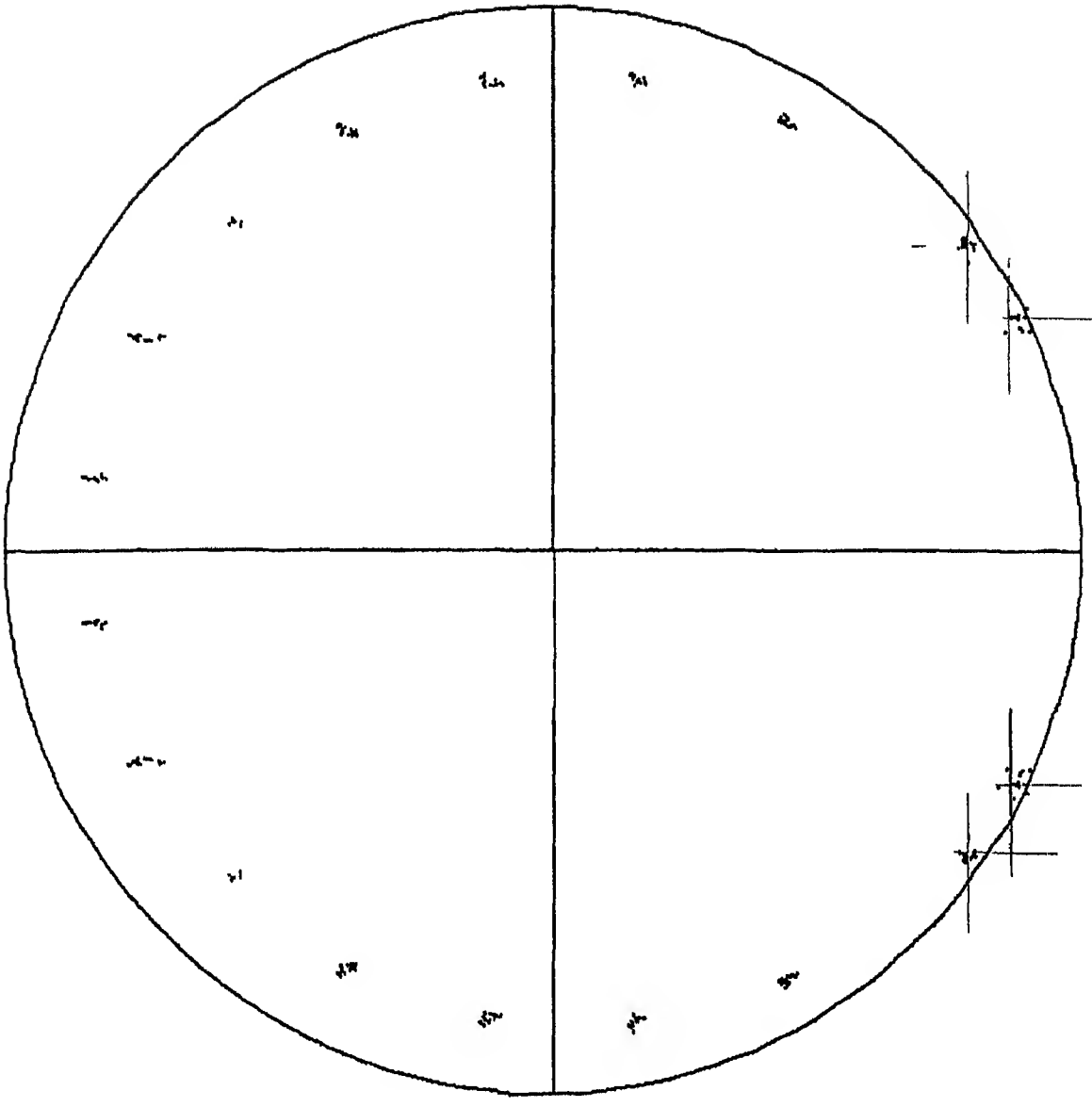

Fig 5.2

Without preprocessing.

SNR = 6 dB, p = 20

Uniform sampling.

the estimate of the parameter vector, $\hat{A}_{MPC}$ The SNR in $y(t_k)$ is 6 dB and the model order, p, used for simulation, is 20 Note that K = L = 50 The signal poles are calculated from the parameter vector, $\hat{A}_{MPC}$, by utilizing equation (4 2 3) The pole estimates for 20 different realizations of $w(t_k)$ are shown in Fig 5 1

5 3 EXAMPLE 2 -

In this example, instead of approximating the given uniformly sampled data by $y(t_k)$, we approxmate it by x[n], as given by equation (4 2 2) The time series for r[n], used in this example, was same as the one used in example 1 This approach, however, differs from the previous one in the sense that, we have not *preprocessed* the given data samples As x[n] is known we can directly use equation (4 6 4) to find the parameter vector, $\hat{A}_{MPC}$ The SNR in x[n] is 6 dB and the model order, p, used for simulation, is 20 Also, L = 50 Note, however, that R'_{ee}, in this case, will be an identity matrix, since, e[n], as given by equation (4 2 2), will represent a white noise sequence The pole estimates, in this case, for 20 different realizations of the white noise sequence, e[n], are plotted in Fig 5 2

5 4 CONCLUSIONS -

The table given below shows the estimates of the bias and variance of the signal parameters for the two cases mentioned above

	SNR = 6 dB p = 20 Example 2		SNR = 6 dB p = 20 Example 1	
	Bias	Variance	Bias	Variance
f_1	$3\ 5819 \times 10^{-3}$	$1\ 3005 \times 10^{-4}$	$8\ 4692 \times 10^{-4}$	$7\ 9826 \times 10^{-5}$
α_1	$4\ 7041 \times 10^{-3}$	$7\ 3334 \times 10^{-4}$	$1\ 0062 \times 10^{4}$	$9\ 2982 \times 10^{-5}$
f_2	$4\ 1137 \times 10^{-3}$	$2\ 5066 \times 10^{-5}$	$3\ 2116 \times 10^{-4}$	$1\ 2982 \times 10^{-5}$
α_2	$6\ 5888 \times 10^{-3}$	$2\ 51815 \times 10^{-5}$	$3\ 9926 \times 10^{-4}$	$1\ 8196 \times 10^{-5}$

Comparing the bias and variance estimates for different signal parameters, we can conclude that the bias and variance estimates in Example 1 are small as compared to those in Example 2

Thus, preprocessing of data points is certainly helpful in increasing the accuracy of the estimates, although, the increase in the computational burden should not be overlooked Lastly, we comment that, the estimation procedures used in Examples 1 and 2 gave similar performance, above a SNR level of 15 dB Thus, for accurate parameter estimates, the additional complexity of preprocesing will be advantageous when the SNR level is below this threshold of 15 dB

REFERENCES

[1] Sircar, P and Sarkar, T K ,"System Identification From Nonuniformly Spaced Signal Measurements", Signal Processing, Vol 14, 1988, pp 253-268

[2] Candy, J V , 'Signal Processing - The Model-Based Approach', McGraw Hill, New York, 1986

[3] Johnson, D H , "The application of spectral estimation methods to bearing estimation", Proc IEEE Vol 70, pp 1018-1028, Sept 1982

[4] Papoulis, A , 'Probability, Random variables and Stochastic Processes", McGraw-Hill, 1984

[5] Schmidt, R O , "A signal subspace approach to multiple emitter location and spectral estimation", Ph D dissertation, Stanford University, 1981

[6] Roy, R and Kailath, T , "ESPRIT - Estimation of Signal Parameters via Rotational Invariance Techniques", IEEE Trans on ASSP, Vol 37, No 7, July 1989

[7] Rabiner, L R & Schafer, R W , 'Digital Processing of Speech Signals', Prentice-Hall, Englewood Cliffs, N J , 1978

[8] Srinivasan, R , "Distributed radar detection theory", IEE Proc"F, Communication, Radar & Signal Processing, Vol 133, Part F, No 1, pp 55-60, Feb 1986

[9] Kumaresan, R and Tufts, D W , "Estimating the parameters of exponentially damped sinusoids and Pole-Zero modelling in noise", IEEE Trans on ASSP, Vol 30, No 6, Dec 1982

[10] Farrier, D R and Wandesforde, A R D , "The estimation of exponentially damped sinusoids with a modified ESPRIT algorithm", Signal Processing 4 Theories and Applications, EURASIP, 1988

[11] Depratterre, E F and Dewilde, P , 'SVD and Signal Processing - Algorithms, Applications and Architectures', Elsevier Science Publications, B V (North - Holland), 1988

[12] Kay, S M , 'Modern Spectral Estimation', Prentice Hall, Englewood Cliffs, New Jersey, pp 422-424

[13] Cadzow, J A and Wu, M M , "Analysis of transient data in noise", IEE Proc , Vol 134, Pt F, No 1, Feb 1987

APPENDIX A

From equation (2 3 9), we can write $(P^T P)^{-1}$ as,

$$(P^T P)^{-1} = \begin{bmatrix} \dfrac{1}{\sum_{k=1}^{K} p_0^2(t_k)} & 0 & 0 \\ 0 & \dfrac{1}{\sum_{k=1}^{K} p_1^2(t_k)} & 0 \\ 0 & 0 & \dfrac{0}{\sum_{k=1}^{K} p_{N-1}^2(t_k)} \end{bmatrix} \quad \text{(A1)}$$

From equation (2 3 6), P^T is given by,

$$P^T = \begin{bmatrix} p_0(t_1) & p_0(t_2) & p_0(t_K) \\ p_1(t_1) & p_1(t_2) & p_1(t_K) \\ p_{N-1}(t_1) & p_{N-1}(t_2) & p_{N-1}(t_K) \end{bmatrix} \quad \text{(A2)}$$

Using equation A1 and A2, $(P^T P)^{-1} P^T$, can be expressed as,

$$(P^T P)^{-1} P^T = \begin{bmatrix} \dfrac{p_0(t_1)}{\sum_{k=1}^{K} p_0^2(t_k)} & \dfrac{p_0(t_2)}{\sum_{k=1}^{K} p_0^2(t_k)} & \dfrac{p_0(t_K)}{\sum_{k=1}^{K} p_0^2(t_k)} \\ \dfrac{p_1(t_1)}{\sum_{k=1}^{K} p_1^2(t_k)} & \dfrac{p_1(t_2)}{\sum_{k=1}^{K} p_1^2(t_k)} & \dfrac{p_1(t_2)}{\sum_{k=1}^{K} p_1^2(t_k)} \\ \dfrac{p_{N-1}(t_1)}{\sum_{k=1}^{K} p_{N-1}^2(t_k)} & \dfrac{p_{N-1}(t_2)}{\sum_{k=1}^{K} p_{N-1}^2(t_k)} & \dfrac{p_{N-1}(t_K)}{\sum_{k=1}^{K} p_{N-1}^2(t_k)} \end{bmatrix} \quad \text{(A3)}$$

Using equations (A3) and (3 2 5), we can write q_{nm}, which is the element at the nth row and mth column of the matrix $P'(P^TP)^{-1}P^T$, as,

$$q_{nm} = \sum_{i=1}^{11} \left(\frac{p_{i-1}[n]\, p_{i-1}[t_m]}{\sum_{k=1}^{k} p_{i-1}[t_k]} \right) \quad , \text{where } p_{i-1}[n] \equiv p_{i-1}[nT] \tag{A4}$$

Note that the *transformation coefficients*, q_{nm}, will be correlated with each other This comes from the fact that the approximating polynomial is a *deterministic* signal Thus, the value of this polynomial at any time instant t_n will be correlated with the value of the same polynomial at different time instant, say t_{n+k} It is apperent, that the amount of correlation will decrease as the lag, k, is increased

APPENDIX B

Let x[n], be a noise corrupted signal, such that, it consists of the signal, s[n], and the noise, w[n], respectively, i e ,

$$x[n] = s[n] + w[n]$$

Let x[0], , x[p], x[p+1], , x[L-1], be the L uniformly spaced samples of the noise corrupted process x[n] Thus, we can write the following equations,

$$\begin{aligned} x[p] &= s[p] + e[p] \\ x[p+1] &= s[p+1] + e[p+1] \end{aligned} \qquad \text{B1}$$

$$x[L-1] = s[L-1] + e[L-1]$$

Equation (B1) can be written in matrix notation as,

$$X = S + W \qquad \text{B2}$$

where,

$$X = [\, x[p] \;\; x[p+1] \;\; \ldots \;\; x[L-1] \,]^T \qquad \text{B3}$$

$$S = [\, s[p] \;\; s[p+1] \;\; \ldots \;\; s[L-1] \,]^T \qquad \text{B4}$$

$$W = [\, w[p] \;\; w[p+1] \;\; \ldots \;\; w[L-1] \,]^T \qquad \text{B5}$$

Let the autocorrelation matrix, R_{xx}, be defined as

$$R_{xx} = E\left(X X^T \right) \qquad \text{B6}$$

Therefore,

$$R_{xx} = E\left[\begin{bmatrix} x[p] \\ x[p+1] \\ \\ \\ x[L-1] \end{bmatrix} \begin{bmatrix} x[p] & x[p+1] & & x[L-1] \end{bmatrix} \right]$$

Therefore,

$$R_{xx} = \begin{bmatrix} E(x[p]x[p]) & E(x[p]x[p+1]) & \cdots & E(x[p]x[L-1]) \\ E(x[p+1]x[p]) & E(x[p+1]x[p+1]) & \cdots & E(x[p+1]x[L-1]) \\ \vdots & \vdots & \ddots & \vdots \\ E(x[L-1]x[p]) & E(x[L-1]x[p+1]) & \cdots & E(x[L-1]x[L-1]) \end{bmatrix} \quad \text{B7}$$

where, the autocorrelation function, r_{xx}, is given by,

$$r_{xx}[n+k, n] = E\left(x^*[n+k]\, x[n] \right) \quad \text{B8}$$

$$= E\left(\left(s^*[n+k] + w^*[n+k] \right) \left(s[n] + w[n] \right) \right)$$

As the crosscorrelation functions,

$$r_{sw}[n+k, n] = E\left(s^*[n+k]\, w[n] \right) \quad \text{B9}$$

$$r_{ws}[n+k, n] = E\left(w^*[n+k]\, s[n] \right) \quad \text{B10}$$

are zero, we get,

$$r_{xx}[n+k, n] = E\left(s^*[n+k]\, s[n] \right) + E\left(w^*[n+k]\, w[n] \right)$$

$$= r_{ss}[n+k, n] + r_{ww}[n+k, n] \quad \text{B11}$$

Substituting, equation B11 in B7, we get,

$$R_{xx} = R_{ss} + R_{ww} \quad \text{B12}$$

If s[n] and w[n] are WSS, then R_{xx} is a symmetric Toeplitx matrix If s[n] and w[n] are locally WSS [4], then R_{xx} is a symmetrix matrix, but not Toeplitz If the processes are nonstationary [4] then, R_{xx} has no special properties, such as symmetricity or Toeplitz structure

APPENDIX C

Let x[n] be a process, which is assumed to be locally WSS [4] Let x[0], , x[L-1], be L uniformly sampled samples of x[n] We define a matrix X, such that,

$$X = \frac{1}{(L-p)^{1/2}} \begin{bmatrix} x[p-1] & x[p-2] & & x[0] \\ x[p] & x[p-1] & & x[1] \\ & & & \\ & & & \\ x[L-2] & x[L-3] & & x[L-p-1] \end{bmatrix} \quad \text{C1}$$

Thus, the p by p autocorrelation matrix, R_{xx}, can be written as,

$$R_{xx} = X^H X \quad \text{C2}$$

Let the singular value decomposition (SVD) of X be given by,

$$X = U \Sigma V^H \quad \text{C3}$$

where,

$$U = [\, u_1 \;\; u_2 \;\; \cdots \;\; u_{L-p} \,], \quad \text{C4}$$

$$V = [\, v_1 \;\; v_2 \;\; \cdots \;\; v_p \,] \quad \text{C5}$$

and

$$\Sigma = \begin{bmatrix} \lambda_1 & 0 & & 0 & 0 \\ 0 & \lambda_2 & 0 & & 0 \\ 0 & & & & \\ & & & & \\ 0 & & & & 0 \\ 0 & 0 & & 0 & \lambda_p \\ 0 & 0 & & & 0 \\ & & & & \\ 0 & 0 & & & 0 \end{bmatrix} \quad \text{C6}$$

Substituting equation (C5) in equation (C6), we get,

$$R_{xx} = (U \Sigma V^H)^H (U \Sigma V^H)$$

$$= V \Sigma^T (U^H U) \Sigma V^H$$

$$= V \Sigma^T \Sigma V^H = V \Gamma V^H$$

where

$$\Gamma = \begin{bmatrix} \lambda_1^2 & 0 & & 0 & 0 \\ 0 & \lambda_2^2 & 0 & & 0 \\ 0 & & & & \\ & & & & \\ 0 & & & 0 & \\ 0 & 0 & & 0 & \lambda_p^2 \end{bmatrix}$$

From equation (C7) we observe that the column and row eigen vectors are same